

Received: 2017.04.07
Accepted: 2017.05.29
Published: 2017.11.25

A Prediction Model with a Combination of Variables for Diagnosis of Lung Cancer

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDE 1 **Xiangsheng Cai**
ABCD 2 **Lu Chen**
ACDF 3 **Tao Kang**
DEF 3 **Yongming Tang**
ABDEF 3 **Teong Lim**
AEF 2 **Meng Xu**
AEFG 1,3,4 **Hongxiang Hui**

1 School of Biotechnology, Southern Medical University, Guangzhou, Guangdong, P.R. China
2 Department of Oncology, The First Affiliated Hospital, Jinan University, Guangzhou, Guangdong, P.R. China
3 Dongguan SMU Metabolic Medicine Limited Company, Dongguan, Guangdong, P.R. China
4 Center for Excellence in Pancreatic Disease, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, U.S.A.

Corresponding Author: Hongxiang Hui, e-mail: Huihongx@gmail.com

Source of support: This work was supported by grants from the National 973 Projects (no. 2011CB504006), the Songshan Lake Science and Technology Fund (no. 2010B025 & no. 2010B026), the Guangdong Science and Technology Fund (no. 2010B090400041), and the Doctoral Program of Higher Education of the Ministry of Education (no. 20104433110014)

Background: Multivariate models with a combination of variables can predict disease more accurately than a single variable employed alone. We developed a logistic regression model with a combination of variables and evaluated its ability to predict lung cancer.

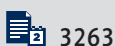
Material/Methods: The exhaled breath from 57 patients with lung cancer and 72 healthy controls without cancer was collected. The VOCs of exhaled breath were examined qualitatively and quantitatively by a novel electronic nose (Z-nose4200 equipment). The VOCs in the 2 groups were compared using the Mann-Whitney *U* test, and the baseline data were compared between the 2 groups using the chi-square test or ANOVA. Variables from VOCs and baseline data were selected by stepwise logistic regression and subjected to a prediction model for the diagnosis of lung cancer as combined factors. The receiver operating characteristic (ROC) curve was used to evaluate the predictive ability of this prediction model.

Results: Nine VOCs in exhaled breath of lung cancer patients differed significantly from those of healthy controls. Four variables – age, hexane, 2,2,4,6,6-pentamethylheptane, and 1,2,6-trimethylnaphthalene – were entered into the prediction model, which could effectively separate the lung cancer samples from the control samples with an accuracy of 82.8%, a sensitivity of 76.0%, and a specificity of 94.0%.

Conclusions: The profile of VOCs in exhaled breath contained distinguishable biomarkers in the patients with lung cancers. The prediction model with 4 variables appears to provide a new technique for lung cancer detection.

MeSH Keywords: **Early Detection of Cancer • Lung Neoplasms • Tumor Microenvironment**

Full-text PDF: <https://www.medscimonit.com/abstract/index/idArt/904738>



Background

Lung cancer is the leading cause of cancer death and the most commonly diagnosed cancer in the world [1]. Early diagnosis of lung cancer is associated with far better survival than diagnosis at a later stage. Five-year survival is expected in 58–73% of patients with stage I lung cancer, while the 5-year survival rate is only 3.5% for later stages of disease. Unfortunately, only 15% of lung cancer patients are diagnosed at stage I, and over half of patients with lung cancer die in the first year after being diagnosed [2].

Screening tests are applied to detect disease in people without any obvious symptoms. Scientists and physicians have looked for many years for a good screening test for lung cancer, such as sputum cytological analysis, chest x-ray, and bronchoscopy, with biopsy currently used in clinical practice [3,4], or tried to develop new biomarkers for lung cancer detection in research [5,6]. However, none of them are ideal and all lack the advantages of being convenient, noninvasive, and effective.

Lung cancers, being located in the lungs, are most likely to release certain characteristic cancer markers directly into the breath. These molecules have attracted research interest because the smell print can be used for lung cancer diagnosis. Volatile organic compounds exhaled in the breath of patients with lung cancer are a focus in this detection [7]. Several previous studies supported the hypothesis that lung cancers release altered breath composition [8–10]. Using gas chromatography/mass spectrometry analysis (GC/MS) on a profile of 9 VOCs, Phillips et al. identified lung cancer patients with 85.1% sensitivity and 80.5% specificity [11]. However, the use of MS for exhaled breath requires expensive equipment and highly skilled analysts. Electronic nose (E-nose) technology is a noninvasive technique, unlike MS, in which array-based sensors look for differences in overall chemical profiles of healthy controls and patients with disease, rather than identifying individual VOC components of exhaled breath. Because of its high sensitivity, the E-nose detection technology is drawing increasing attention in diagnosis, detection of renal dysfunction, chronic obstructive pulmonary disease, asthma, pneumonia, tuberculosis, and lung cancer [12–18]. Using an E-nose for lung cancer screening may be a radical innovation in diagnosis of this cancer and other pulmonary diseases.

The applications of tiny amounts of VOCs in exhaled breath are drawing more and more attention in the development of diagnoses, such as tuberculosis, diabetes, breast cancer, lung cancer, colon cancer, and prostate cancer [19–23]. It has been 30 years since Gordon pioneered use of VOCs to diagnose lung cancer. However, due to the use of different biomarkers and detection equipment, there is no consensus regarding acceptable biomarkers and research. Use of the electronic nose

(E-nose) has only recently gained attention for its sensitive detection ability. It has been widely used in cocaine detection, environmental pollution monitoring, and biomarker detection. Phillips reported that diagnosing lung cancer and other respiratory cancers by use of the E-nose provides improved sensitivity and specificity [20,21]. Use of an E-nose enhances the early diagnosis of lung cancer by detecting biomarkers in exhaled breath and establishing the “finger print” data of patients with lung cancer.

We used a novel electronic nose (zNOSE4200), which contains a surface acoustic wave quartz microbalance sensor, to analyze “smell prints” of exhaled gases and to select a set of VOCs as lung cancer biomarkers. We created a cancer prediction model using variables significantly related to the presence of lung cancer. The discriminatory power of the model was tested to validate the potential utility of “smell print” signatures for identifying lung cancer.

Material and Methods

Subjects

A total of 57 patients with lung cancer who visited the First Affiliated Hospital of Jinan University during January 2013 to January 2014 were recruited. Among these 57 lung cancer patients, 53 had non-small-cell lung cancer (NSCLC) and 4 had small-cell lung cancer, all confirmed by histopathological examination and diagnosed as the original lung cancers. Their ages were 37–70 years, and there were 36 males and 21 females. A total of 72 healthy people from Guangzhou were recruited as the control group, with ages 30–58 years. None of the patients or healthy controls had any other chronic diseases (e.g., diabetes, chronic renal insufficiency, chronic obstructive pulmonary disease, pulmonary tuberculosis, asthma, and upper respiratory infection). Patients who had received other treatments (e.g., chemotherapy and surgical operation) after diagnosis of LC were also excluded. This study was approved by the Institutional Ethics Review Board of the First Affiliated Hospital of Jinan University and every subject signed a written informed consent.

Collection and preparation of the breath samples

On the day prior to the test, certain foods with strong smells, such as onions, leeks, garlic, cabbage, and pickled cabbage or beans, were avoided. Prior to the test, the subjects were asked to stop smoking and chewing gum. Subjects were asked to fast for at least 12 h before testing and to only consume water. Breath sample collections were performed by having the subjects inhale, hold their breath momentarily, and then exhale into a 1-liter Devex bag. After collection, each sample bag

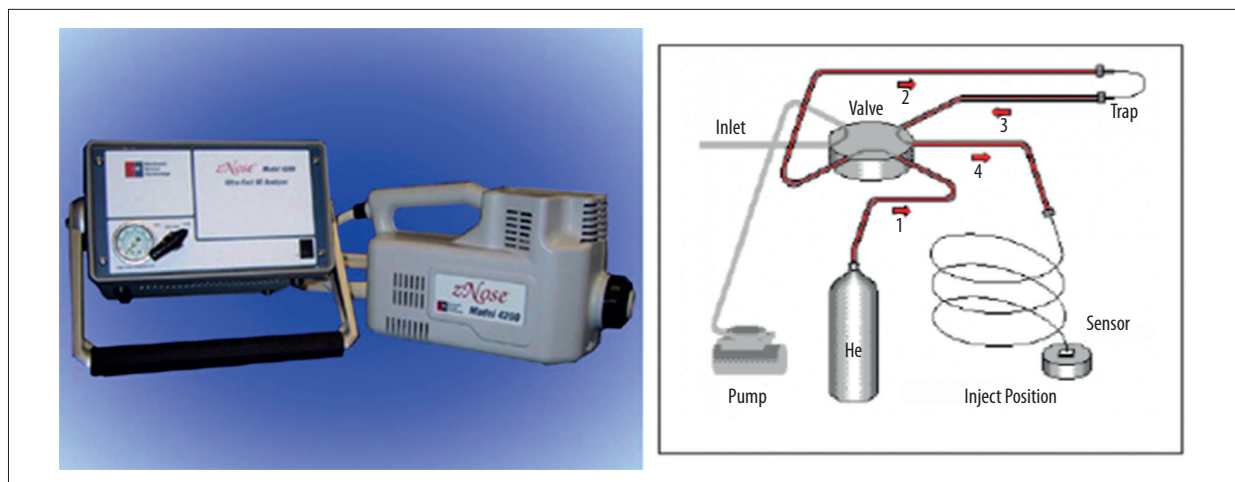


Figure 1. Schematic diagram of z-NOSE-4200. The GC sensor is based on a 6-port valve and oven, a pre-concentrating trap, a short GC column, and a surface acoustic wave detector.

was immediately attached to the sample inlet of the E-nose for analysis.

VOCs analysis

The E-nose device (zNOSE4200, Electronic Sensor Technology, Newbury Park, CA, USA), which contains high-speed gas chromatograph (GC) sensors, was used in our studies. The GC sensor is based on a 6-port valve and oven, a pre-concentrating trap, a short GC column, and a surface acoustic wave quartz microbalance detector (Figure 1). The sample was induced through a heated sample loop at 200°C driven by an internal pump at a flow rate of 30 ccm. The time course for a single sample analysis is 60 s. Helium gas was used as the carrier medium with a flow rate of 1.8 mL min⁻¹. This equipment can obtain a sensitivity below ppb level with 5% RSD (relative standard deviation). The unique advantages of this device are its small in size and light weight, allowing it to be conveniently carried and set up in a clinical room for use (Figure 1).

Retention time was used to identify the alkane molecules in VOC by comparing the retention time of unknown molecules to those of a series of C₆–C₁₄ alkane standards. Prior to the analysis of VOC, the device was calibrated with the standards (1.25 µl each), including hexane (C6), heptane (C7), octane (C8), nonane (C9), decane (C10), undecane (C11), dodecyl (C12), tridecane (C13), and tetradecane (C14). The retention time of each alkane sample was converted to Kovats Indices (KI), a standard index for retention time in chromatography. The chemical structure and the name of each VOC was determined according to Kovatz Index.

To quantify the concentration of detected molecules, the device was also calibrated with a series of C₆–C₂₀ alkane standards (0.2 µl each with known concentration), including hexane

(C6), heptane (C7), octane (C8), nonane (C9), decane (C10), undecane (C11), dodecyl (C12), tridecane (C13), tetradecane (C14), hexadecane (C16), octadecyl (C18), and eicosane (C20). The area of each peak and its representing amount of standard were used to estimate the abundance of peaks of interest identified during VOC analysis.

Biomarkers identification

Compounds in VOC were identified by comparing their retention times with those of the standards (Figure 2). The mass spectrums were confirmed by Kovats Indices according to the mass spectrometry libraries NIST 05 and NIST 05s. The abundance of each sample was calculated by comparing its peak area with those of the standards (Figure 3). All the VOCs with a peak area value higher than 200 were detected. The compounds existing in the breath but not existing in the room air and the compounds with higher abundance in breath than in room air were considered as the endogenous VOCs.

Candidate biomarkers of lung cancer were initially identified by comparing the abundance of endogenous VOCs between the primary lung cancer group and the control group. The Mann-Whitney *U* test was used for the mathematic analysis of correlation.

Validation of prediction model

To assess the performance of the prediction model, we used a validation cohort of 118 subjects (untreated primary lung cancer and cancer-free controls) recruited at the Inpatient Department of Oncology of the First Affiliated Hospital of Jinan University. These subjects underwent physical examination at Outpatient Department and the results showed they were suspected to have LC.

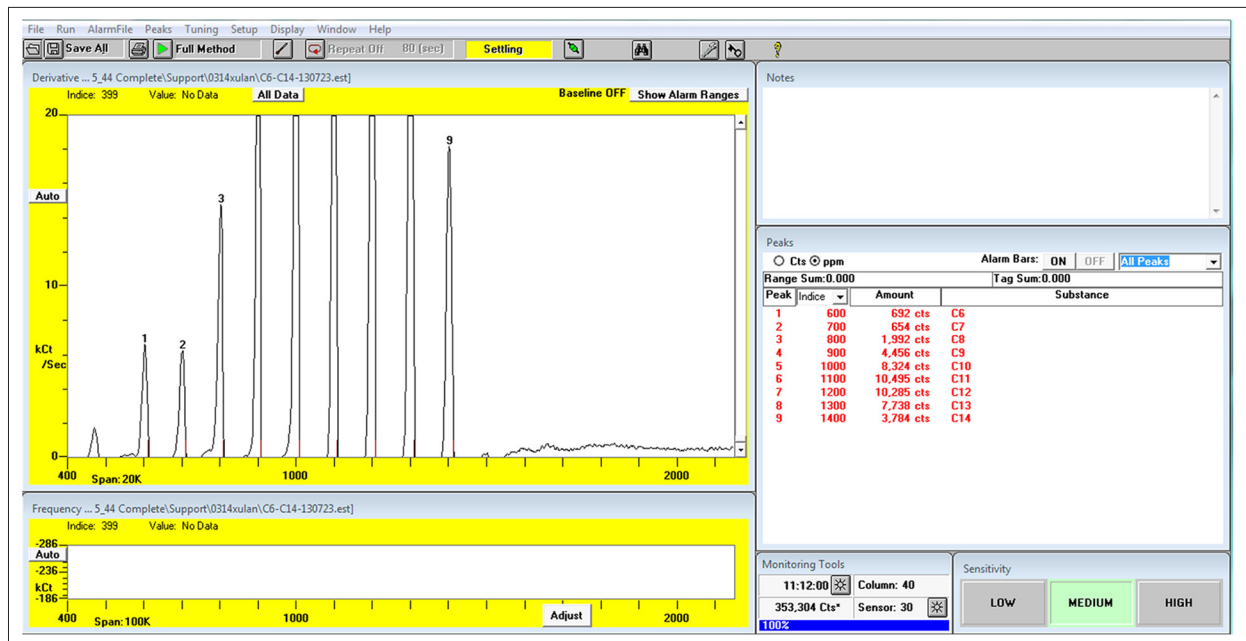


Figure 2. The output of the Z-nose 4200 after calibration by the standard solution (C_6-C_{14}).

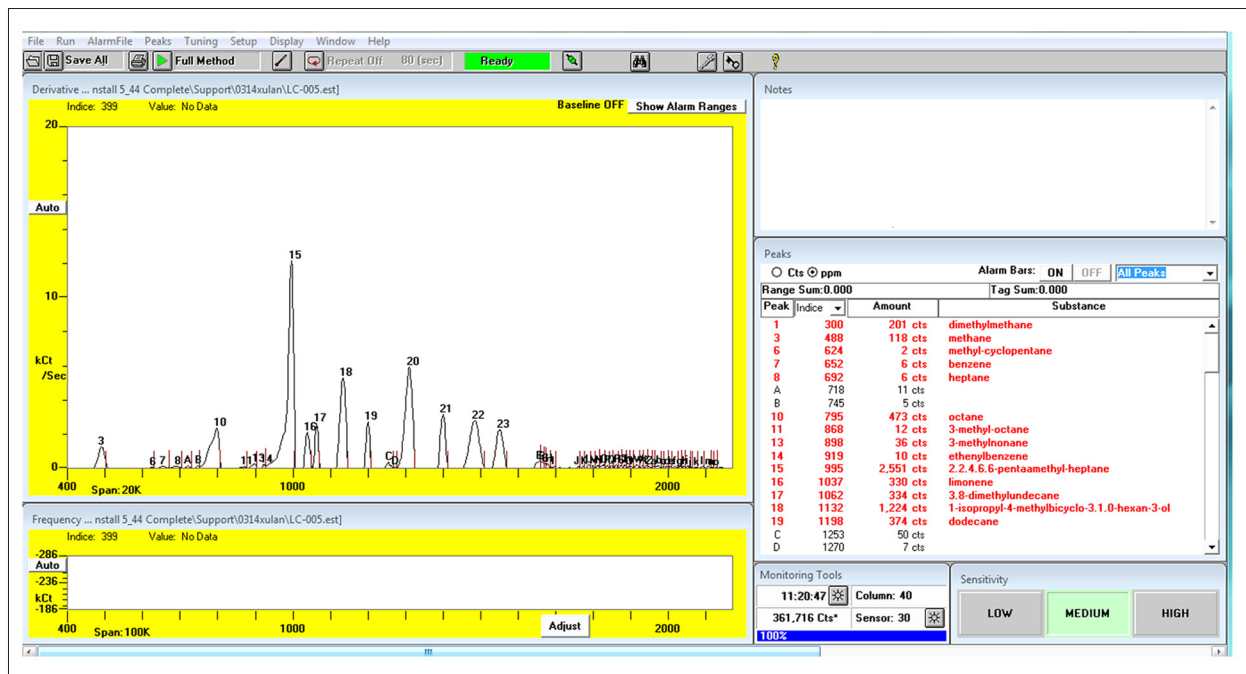


Figure 3. The result of exhaled gas detected by Z-nose 4200 after qualitative and quantitative calibration.

Statistical analysis

Statistical analyses were performed with SPSS (IBM SPSS Statistics 20) and MATLAB (R2008b). The statistical results are expressed as means (or medians) \pm SD and versus the t test or the Mann-Whitney non-parametric test. Categorical values are described by counts and proportions referencing the chi-squared test.

Because the normalized or corrected count rates did not fit a normal distribution for most masses, we screened for significant differences in different data sets using the two-tailed non-parametric Mann-Whitney U test, which is based on the overall ranking order of individual values from 2 data sets. It has a similar efficiency as the t test when testing samples from presumably identical distributions for shifts in location. Hence, it is not limited to normal distribution data. The distribution of

Table 1. The composition of the subject panel.

| Characters | Groups | | χ^2/t | P | |
|---------------|-----------------|-------------|------------|-------|--------|
| | Lung cancer (n) | Control (n) | | | |
| Sex | Male | 36 | 39 | 1.057 | 0.304 |
| | Female | 21 | 33 | | |
| Age | $\bar{x}\pm s$ | 58.34±14.35 | 46.38±9.72 | 6.022 | <0.001 |
| Smoker status | Smoker | 22 | 25 | 0.206 | 0.649 |
| | Nonsmoker | 35 | 47 | | |

its large sample test statistic z is approximately standard normal; an absolute value of z exceeding 1.96 represents a significant difference between the 2 data sets at a level of 5%.

Variables significantly related to the presence of lung cancer were input into the logistic regression (LR) model for analysis. Receiver operating characteristic (ROC) curves for 4 factors, and the LR model were tested for usefulness in discerning lung cancer based on their areas under the curve (AUC).

Results

Human subjects

No adverse effects were observed in any subject as a result of donating a breath sample. These observations confirmed that this assay is noninvasive and safe. The composition of the subject panel and the characters are summarized in Table 1. Although smoking status and sex are assumed to affect cancer incidence, our subject composition revealed no significant difference in smoking status ($\chi^2=0.206$, $P=0.649$), and sex ($\chi^2=1.057$, $P=0.304$), but age was significantly different between the 2 groups ($t=6.022$, $P<0.001$; Table 1).

Selection of candidate VOC biomarkers for lung cancer

The VOCs of exhaled breath from 129 subjects were analyzed. Twenty-three promising endogenous VOCs with differences between the LC group and control group were selected as the diagnostic candidates (Table 2).

Further identification was conducted by Mann-Whitney U test in a non-parametric math model. The z values of the U tests are frequently lower than 1.96 in the patients with lung cancer ($P=0.05$). Nine markers exhibited promising results with significant differences in abundance: dimethylmethane [$Z=-2.426$, $P=0.015$], ethanol [$Z=-2.470$, $P=0.014$], methane [$Z=-1.989$, $P=0.047$], hexane [$Z=-2.321$, $P=0.020$], 2.2.4.6.6-pentamethylheptane [$Z=-4.543$, $P=0.001$],

Table 2. Endogenous VOCs tested as candidate biomarkers for lung cancer (n=23).

| Name | Kovats index (KI) |
|--|-------------------|
| Dimethylmethane | 304 |
| Ethanol | 460 |
| Methane | 483 |
| Isoprene | 520 |
| Hexane | 600 |
| Methylcyclopentane | 627 |
| Benzene | 654 |
| Heptane | 700 |
| 2-methylheptane | 763 |
| Octane | 800 |
| 3-methyloctane | 871 |
| 1.4-dimethylbenzene | 877 |
| 3-methylnonane | 900 |
| Ethylbenzene | 915 |
| 2.2.4.6.6-pentamethylheptane | 997 |
| Limonene | 1038 |
| 2,5,5-trimethyl-2,6-heptadien-4-one | 1063 |
| 1-isopropyl-4-methylbicyclo[3.1.0]hexan-3-ol | 1133 |
| Dodecane | 1200 |
| Tridecane | 1300 |
| Tetradecane | 1400 |
| 2-phenylpropylbutyrate | 1484 |
| 1,2,6-trimethylnaphthalene | 1552 |

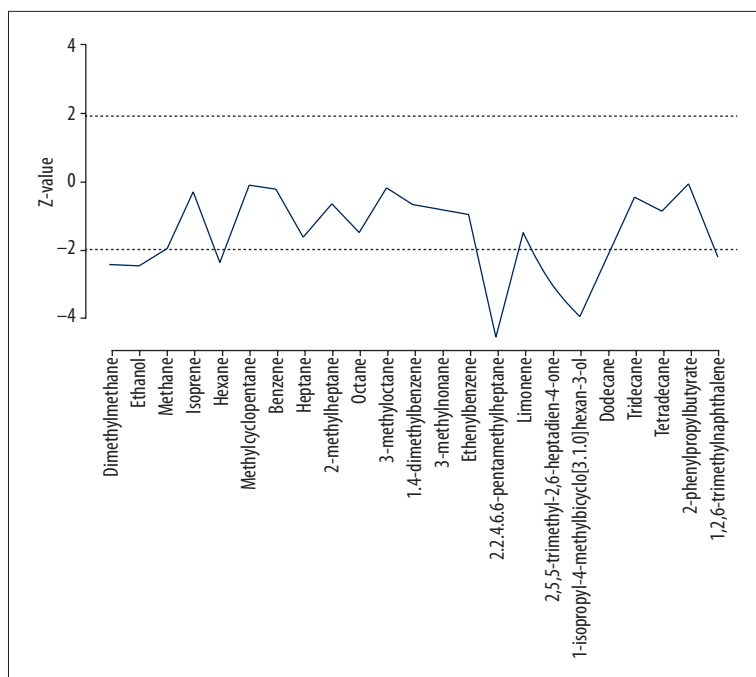


Figure 4. Non-parametric Mann-Whitney *U* test comparing Z values of 23 VOCs in the lung cancer and healthy control groups. Dashed line: significance of $Z=\pm 1.96$. The VOCs (dimethylmethane [$Z=-2.426$, $P=0.015$], ethanol [$Z=-2.470$, $P=0.014$], methane [$Z=-1.989$, $P=0.047$], hexane [$Z=-2.321$, $P=0.020$], 2,2,4,6,6-pentamethylheptane [$Z=-4.543$, $P=0.001$], 2,5,5-trimethyl-2,6-heptadien-4-one [$Z=-2.926$, $P=0.003$], 1-isopropyl-4-methylbicyclo [3.1.0] hexan-3-ol [$Z=-3.904$, $P=0.001$], dodecane [$Z=-2.137$, $P=0.033$], and 1,2,6-trimethylnaphthalen [$Z=-2.241$, $P=0.025$]). Z values that exceeded the dashed line significantly differ between the lung cancer and healthy control groups.

2,5,5-trimethyl-2,6-heptadien-4-one [$Z=-2.926$, $P=0.003$], 1-isopropyl-4-methylbicyclo [3.1.0] hexan-3-ol [$Z=-3.904$, $P=0.001$], dodecane [$Z=-2.137$, $P=0.033$], 1,2,6-trimethylnaphthalene [$Z=-2.241$, $P=0.025$] (Figure 4).

The mean rank of VOCs in the lung cancer group revealed that dimethyl methane [65.50, 70.87], ethanol [60.23, 76.88], methane [65.29, 71.11], hexane [60.79, 76.25], 2,2,4,6,6-pentamethylheptane[53.69, 84.34], 1-isopropyl-4-methylbicyclo[3.1.0]hexan-3-ol [55.71, 82.04], and dodecane [61.27, 75.68] in the lung cancer group were higher than those in healthy people, and the concentrations of 2,5,5-trimethyl-2,6-heptadien-4-one [77.22, 57.48] and 1,2,6-trimethyl naphthalene [75.06, 59.94] in lung cancer group were lower than in the healthy group (Table 3).

Mathematics analysis model of lung cancer detection

The Mathematics logistic regression model was applied for variables from the baseline data (sex, age, smoking status) and the parameters of 9 VOCs by the Stepwise Regression Analysis. We found 4 parameters: age (X_1), hexane (X_2), 2,2,4,6,6-pentamethyl heptane (X_3) and 1,2,6-trimethyl naphthalene (X_4) and their combination were significant in the logistic regression model (Table 4). We developed a formula of $P=1/[1+e^{(-9.005+102*X_1+0.011*X_2+0.022*X_3-0.517*X_4)}]$, which is able to predict the lung cancer probability in this assay with statistical significance ($\chi^2=59.360$, $P=0.001$). Among these 4, 3 parameters: age, hexane, 2,2,4,6,6-pentamethyl heptane are risk factors, while the last one: 1,2,6-trimethyl naphthalene is a protective factor. The statistical analysis results indicated that the diagnostic prediction by this formula was able to distinguish

the breath sample of the lung cancer patients from those of the normal subjects with 82.8% accuracy, 76% sensitivity and 94% specificity (Table 5).

ROC analysis was conducted based on results of definitive diagnoses. Figure 5 displays the validation ROC curves. The accuracy of the prediction derived from the formula that was input with 4 parameters seems better than each of the ROC curves of the LR model. The sensitivity and specificity of this LR model were 76% and 94%, respectively. The AUC (0.878) from the formula was 0.878, which was higher than the AUCs for age (0.774), hexane (0.622), 2,2,4,6,6-pentamethyl heptane (0.726), and 1,2,6-trimethyl naphthalene (0.381). These comparisons indicate that the diagnostic prediction ability of the combination of 4 parameters is better than that of each single one.

Discussion

Early detection will be extremely valuable in the battle against lung cancer. Unfortunately, a noninvasive, convenient, and effective screening for lung cancer is still not available. Breath analysis should be an ideal detection approach for such a screening due to its non-invasiveness, simplicity, and low cost. Early diagnosis and real-time monitoring of lung cancer by detection of certain molecules in exhaled breath has become an important research focus. Recent publications have shown that the difference between breath samples from patients with lung cancer and those from healthy subjects can be distinguished by use of an E-nose [12,15,19], which is a gas chromatography-based instrument

Table 3. Mann-Whitney U test compares Z values of 23 VOCs in the lung cancer and healthy control groups.

| Name | Kovarts index (KI) | Mean rank | | U | Z | P |
|--|--------------------|-----------|-------------|--------|--------|-------|
| | | Control | Lung cancer | | | |
| Dimethylmethane | 304 | 65.50 | 70.87 | 2088.0 | -2.426 | 0.015 |
| Ethanol | 460 | 60.23 | 76.87 | 1708.5 | -2.470 | 0.014 |
| Methane | 483 | 65.29 | 71.11 | 2073.0 | -1.989 | 0.047 |
| Isoprene | 520 | 67.24 | 68.88 | 2213.0 | -0.271 | 0.786 |
| Hexane | 600 | 60.79 | 76.25 | 1749.0 | -2.321 | 0.020 |
| Methylcyclopentane | 627 | 68.35 | 67.61 | 2243.0 | -0.113 | 0.910 |
| Benzene | 654 | 67.17 | 68.93 | 2208.5 | -0.266 | 0.790 |
| Heptane | 700 | 73.17 | 62.11 | 1896.0 | -1.642 | 0.101 |
| 2-methylheptane | 763 | 69.49 | 66.26 | 2160.5 | -0.636 | 0.525 |
| Octane | 800 | 72.53 | 62.82 | 1942.0 | -1.438 | 0.150 |
| 3-methyloctane | 871 | 67.51 | 68.57 | 2232.5 | -0.157 | 0.875 |
| 1,4-dimethylbenzene | 877 | 69.65 | 66.14 | 2149.0 | -0.662 | 0.508 |
| 3-methylnonane | 900 | 65.40 | 70.93 | 2081.0 | -0.826 | 0.409 |
| Ethenylbenzene | 915 | 65.04 | 71.39 | 2055.0 | -0.949 | 0.343 |
| 2.2.4.6.6-pentamethylheptane | 997 | 53.69 | 84.34 | 1238.0 | -4.543 | 0.000 |
| Limonene | 1038 | 72.68 | 62.66 | 1931.0 | -1.486 | 0.137 |
| 2,5,5-trimethyl-2,6-heptadien-4-one | 1063 | 77.22 | 57.48 | 1604.5 | -2.926 | 0.003 |
| 1-isopropyl-4-methylbicyclo[3.1.0]hexan-3-ol | 1133 | 55.71 | 82.04 | 1383.0 | -3.904 | 0.000 |
| Dodecane | 1200 | 61.27 | 75.68 | 1783.5 | -2.137 | 0.033 |
| Tridecane | 1300 | 69.31 | 66.52 | 2174.0 | -0.416 | 0.677 |
| Tetradecane | 1400 | 65.31 | 71.07 | 2074.5 | -0.853 | 0.393 |
| 2-phenylpropylbutyrate | 1484 | 67.97 | 68.03 | 2266.0 | -0.009 | 0.993 |
| 1,2,6-trimethylnaphthalene | 1552 | 75.06 | 59.94 | 1760.0 | -2.241 | 0.025 |

Table 4. Variables in the Equation.

| Factor | B | SE | Wald | P | OR | 95% C.I. for OR | |
|------------------------------|--------|--------|--------|--------|-------|-----------------|-------|
| | | | | | | Lower | Upper |
| Age | 0.102 | 0.022 | 18.803 | 0.001 | 1.105 | 1.058 | 1.157 |
| Hexane | 0.011 | 0.0001 | 6.276 | 0.011 | 1.002 | 1.000 | 1.001 |
| 2.2.4.6.6-pentamethylheptane | 0.022 | 0.005 | 12.757 | <0.001 | 1.022 | 1.010 | 1.031 |
| 1,2,6-trimethylnaphthalene | -0.517 | 0.203 | 6.664 | 0.010 | 0.594 | 0.403 | 0.882 |
| Constants | -9.005 | 1.884 | 22.777 | <0.001 | 0.001 | | |

Table 5. Predicted probability and diagnostic values of significant variables.

| Factor | Diagnosis cut-off point | Sensitivity (%) | Specificity (%) | Area under the curve (AUC) | 95% C.I. for AUC | P |
|------------------------------|-------------------------|-----------------|-----------------|----------------------------|------------------|--------|
| Age | 57.4 | 0.58 | 0.95 | 0.774 | 0.685~0.854 | 0.001 |
| Hexane | 480.96 | 0.68 | 0.57 | 0.622 | 0.529~0.718 | 0.015 |
| 2.2.4.6.6-pentamethylheptane | 214.11 | 0.59 | 0.84 | 0.726 | 0.634~0.816 | <0.001 |
| 1,2,6-trimethylnaphthalene | 0.034 | 0.94 | 0.15 | 0.381 | 0.281~0.478 | 0.017 |
| Predicted probability | 0.61 | 0.76 | 0.94 | 0.878 | 0.810~0.940 | <0.001 |

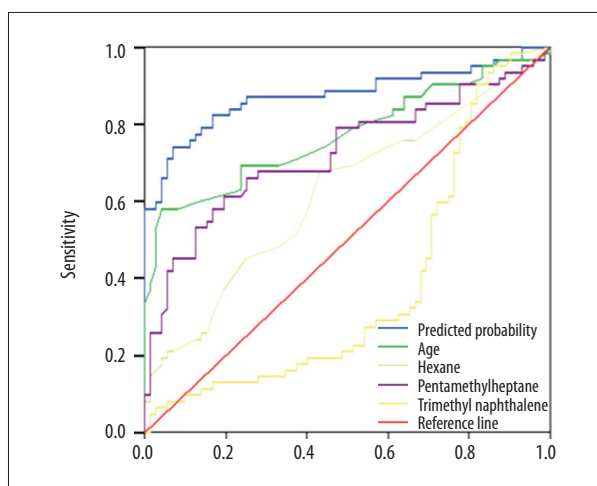


Figure 5. ROC curves of predicted probability and diagnostic values of significant variables. The AUC of the LR model was 0.878, which was higher than AUCs for age (0.774), hexane (0.622), 2.2.4.6.6-pentamethylheptane (0.726), and 1.2.6-trimethyl naphthalene (0.381).

Although there are different types of E-nose sensors, a more sensitive instrument with tiny alternation is expected to detect the VOCs. In our study, we used a novel E-nose (Znose4200), which had a surface acoustic wave quartz microbalance sensor, to analyze the VOCs of exhaled breath. The SAW resonator detector oscillates at 500 MHz. This sensor can accurately measure relative changes at frequencies as low as 1 Hz. Therefore, the SAW gas sensor exhibits an excellent threshold detection limit below *ppb*.

The exhaled breath contains tiny amounts of VOCs, as well as other major components such as nitrogen, oxygen, carbon dioxide, and water. Up to 3450 tiny components can be detected, with a concentration in *ppm* or *ppt* levels [24,25]. In subjects with inflammation, oxygen stress, or cancers, the altered metabolism will release altered molecules into the blood and ultimately release them into the exhaled breath [25]. The altered concentrations of certain biomarkers reflect the altered

status of physiological or pathological status. For example, the strong smell of acetone (rotten apple) in type I diabetic patients indicates an acetone toxic status [26].

Ideally, the collected exhaled breath should be identical to the alveolar air sample. To minimize the possible dilution, we collected the last half fraction of the 4th exhaled breath after 3 deep breaths. We believe this minimizes the possible dilution and prevents contamination from environmental air [27].

According to recent reports, the lung cancer-specific VOCs can be divided into 7 categories: hydrocarbons (alkanes, branched alkanes, branched olefins), alcohols, aldehydes, ketones, esters, nitriles, and aromatic compounds [7]. The metabolic mechanism of VOCs in breath is still unclear. The general belief is that the VOCs are not solely from the tumor tissue, but also from a series of metabolic disorders [28,29]. In the present study, the 9 VOCs selected as the lung cancer markers came from 4 families: hydrocarbons (dimethylmethane, methane, hexane, dodecane and 2.2.4.6.6-pentamethylheptane), alcohols (ethanol and 1-isopropyl-4-methylbicyclo [3.1.0] hexan-3-ol), ketones (2.5.5-trimethyl-2.6-heptadien-4-one), and aromatic compounds (1.2.6-trimethylnaphthalene). Hydrocarbons are produced in lipid peroxidation of polyunsaturated fatty acids by reactive oxygen, which is elevated with tumor progression [30,31]. Most of the alcohols are generated from food in the gastrointestinal tract, and a small portion are from hydrocarbon metabolism. Ketones are generated during peroxidation of polyunsaturated fatty acids and can be detected in exhaled breath [7]. Aromatic compounds are considered to be exogenous pollutants derived from air pollution, radiation exposure, cigarette smoke, and alcohol drinking [32].

Previous studies reported that the abundance of decane, 2-methyl pentane, 1-propanol, 2-butanone, 2-ethyl hexanol, benzene, and toluene were higher in the lung cancer group than in the control group [33–35]. The present study, for the first time, found that the concentrations of dimethyl methane, ethanol, methane, hexane, 2.2.4.6.6-pentamethylheptane, 1-isopropyl-4-methylbicyclo [3.1.0] hexan-3-ol, and dodecane were

higher in the lung cancer group than in the control group; however, the abundance of 2,5,5-trimethyl-2,6-heptadien-4-one and 1,2,6-trimethyl naphthalene in the lung cancer group was lower than in the healthy group. The mechanism remains unclear.

For many years, due to the use of various different detection equipment and gas collection procedures, the results reported in the literature are inconsistent. Kischkel [36] reported that the concentrations of acetone in the exhaled breath of patients with lung cancer are higher than in healthy people, but Bajtarevic [33] drew the opposite conclusion. Bajtarevic [33] reported that the concentration of isoprene in the exhaled breath of lung cancer patients was lower than in healthy people, while Poli reported the opposite result [34]. In addition, the sensitivity and specificity from numerous studies vary. Phillips found 30 kinds of VOCs in expiratory samples by using GC-MS detection with 84.5% sensitivity and 81% specificity (193 samples from lung cancer patients and 211 samples from disease-free persons) [21], while Machado separated lung cancer patients from control samples with 71.4% sensitivity and 91.9% specificity by electronic nose detection (28 samples from lung cancer patients and 109 samples from disease-free persons) [16]. Mazzone distinguished lung cancer patients from control subjects with 73.3% sensitivity and 72.4% specificity by using a colorimetric sensor array (49 samples from lung cancer patients and 21 samples from disease-free persons) [35]. Mazzone indicated that smoking, age, and the stage of the disease did not affect their results [35].

The specificity of the biomarkers of lung cancers is critical during biomarker selection. The markers of acetone and ethanol were excluded as they are associated with other diseases such as diabetes [37] and ketosis [38]. In addition, the concentration levels of acetone varied widely from study to study. For example, Kischkel et al. [36] observed an increase in acetone concentration in the breath of LC subjects relative to healthy controls, but Bajtarevic et al. [33] reported the opposite trend. The concentration levels of ethanol were reported to vary in different hepatocarcinoma cell lines *in vitro* [39].

References:

- Jemal A: Global cancer statistics. *Cancer J Clin*, 2011; 61(2): 69–90; Erratum in: *Cancer J Clin*, 2011; 61(2): 134
- Tarver T: Cancer facts & figures 2012. American Cancer Society (ACS). *Journal of Consumer Health on the Internet*, 2012; 16: 366–67
- Todd J, Mcgrath EE: Chest X-ray mass in a patient with lung cancer! *QJM*, 2011; 104: 903–4
- Wang YXJ, Lo GG, Yuan J, Larson PEZ et al: Magnetic resonance imaging for lung cancer screen. *J Thorac Dis*, 2014; 6: 1340–48
- Huang Q, Li L, Lin Z et al: Identification of preferentially expressed antigen of melanoma as a potential tumor suppressor in lung adenocarcinoma. *Med Sci Monit*, 2016; 22: 1837–42
- Zhou B, Xu H, Ni K et al: Expression of chemokine XCL2 and CX3CL1 in lung cancer. *Med Sci Monit*, 2016; 22: 1560–65
- Hakim M, Broza YY, Barash O, Peled N et al: Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem Rev*, 2015; 112: 5949–66
- Wang Y, Hu Y, Wang D, Yu K et al: The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath, tissues and cell lines. *Cancer Biomark*, 2012; 11: 129–37
- Peng G, Tisch U, Adams O et al: Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nat Nanotechnol*, 2009; 4: 669–73
- Phillips M, Altorki N, Austin JH et al: Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomark*, 2007; 3: 95–109
- Phillips M, Cataneo RN, Cummin AR et al: Detection of lung cancer with volatile markers in the breath. *Chest*, 2003; 123: 2115–23

The present study achieved prediction of lung cancer versus controls with 82.8% accuracy, 76% sensitivity, and 94% specificity. Although variable smoking habit was expected to be a confounding factor in our detection system, based on the published data [16,35], it did not change the discriminative power of VOC in our tests because smoking status was not included as a factor into our predicted model. Unlike others prediction models which compare the difference of single VOCs, our model combined multiple parameters as a comprehensive evaluation, ensuring better authenticity and reliability of the results.

Despite these achievements, there are several limitations in the present study. A comprehensive and systematic analysis should be conducted to explore the differences between subject characteristics (e.g., sex, smoking status, family history of lung cancer, occupational exposure, and chronic lung disease history) and VOCs in the exhaled breath.

Conclusions

We demonstrated the feasibility of detection of certain markers in VOCs of exhaled breath using a novel electronic nose (zNOSE4200) containing a SAW gas sensor as a relatively convenient and noninvasive test in patients with suspected lung cancer. A prediction model combined with variables of, age, hexane, and 2,2,4,6,6-pentamethyl heptane, and 1,2,6-trimethyl naphthalene concentration could be used to effectively predict the presence of lung cancer. Further study is needed to optimize the accuracy of the prediction model by incorporating a clinical phenotype and to develop a promising clinical detection method of lung cancer.

Conflict of Interest

None.

12. Hockstein NG, Thaler ER, Torigian D et al: Diagnosis of pneumonia with an electronic nose: Correlation of vapor signature with chest computed tomography scan findings. *Laryngoscope*, 2004; 114: 1701–5.
13. Dragonieri S, Annema JT, Schot R, et al: An electronic nose in the discrimination of patients with non-small cell lung cancer and COPD. *Lung Cancer*, 2009; 64: 166–70
14. Fens N1, Zwinderman AH, van der Schee MP et al: Exhaled breath profiling enables discrimination of chronic obstructive pulmonary disease and asthma. *Am J Respir Crit Care Med*, 2009; 180: 1076–82
15. Dragonieri S, Schot R, Mertens BJ et al: An electronic nose in the discrimination of patients with asthma and controls. *J Allergy Clin Immunol*, 2007; 120: 856–62
16. Machado RF, Laskowski D, Deffenderfer O et al: Detection of lung cancer by sensor array analyses of exhaled breath. *Am J Respir Crit Care Med*, 2005; 171: 1286–91
17. Phillips M, Basa-Dalay V, Blais J et al: Point-of-care breath test for biomarkers of active pulmonary tuberculosis. *Tuberculosis*, 2012; 92: 314–20
18. Voss A, Baier V, Reisch R et al: Smelling renal dysfunction via electronic nose. *Ann Biomed Eng*, 2005; 33: 656–60
19. Peng G, Hakim M, Broza YY et al: Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *Br J Cancer*, 2010; 103: 542–51
20. Phillips M, Cataneo RN, Saunders C et al: Volatile biomarkers in the breath of women with breast cancer. *J Breath Res*, 2010; 4: 026003
21. Phillips M, Altorki N, Austin JH et al: Detection of lung cancer using weighted digital analysis of breath biomarkers. *Clin Chim Acta*, 2008; 393: 76–84
22. Phillips M, Cataneo RN, Cheema T, Greenberg J: Increased breath biomarkers of oxidative stress in diabetes mellitus. *Clin Chim Acta*, 2004; 344: 189–94
23. Gordon SM, Szidon JP, Krotoszynski BK et al: Volatile organic compounds in exhaled air from patients with lung cancer. *Clin Chem*, 1985; 31: 1278–82
24. Phillips M, Herrera J, Krishnan S et al: Variation in volatile organic compounds in the breath of normal humans. *J Chromatogr B Biomed Sci Appl*, 1999; 729(1–2): 75–88
25. Boots AW, van Berkel JJ, Dallinga JW et al: The versatile use of exhaled volatile organic compounds in human health and disease. *J Breath Res*, 2012; 6: 027108
26. Libardoni M, Stevens PT, Waite JH, Sacks R: Analysis of human breath samples with a multi-bed sorption trap and comprehensive two-dimensional gas chromatography (GCxGC). *J Chromatogr B Analyt Technol Biomed Life Sci*, 2006; 842(1): 13–21
27. Miekisch W, Schubert JK, Noeldgeschomburg GF: Diagnostic potential of breath analysis – focus on volatile organic compounds. *Clin Chim Acta*, 2004; 347: 25–39
28. Poli D, Goldoni M, Caglieri A et al: Breath analysis in non small cell lung cancer patients after surgical tumour resection. *Acta Biomed*, 2008; 79(Suppl 1): 64–72
29. Amann A, Corradi M, Mazzone P, Mutti A: Lung cancer biomarkers in exhaled breath. *Expert Rev Mol Diagn*, 2011; 11: 207–17
30. Phillips M, Cataneo RN, Greenberg J et al: Effect of oxygen on breath markers of oxidative stress. *Eur Respir J*, 2003; 21(1): 48–51; Erratum in: *Eur Respir J*, 2003; 21(5): 911
31. Mitsui T, Kondo T: Inadequacy of theoretical basis of breath methylated alkane contour for assessing oxidative stress. *Clin Chim Acta*, 2003; 333: 91; author reply 93–94
32. Stone BG, Besse TJ, Duane WC et al: Effect of regulating cholesterol biosynthesis on breath isoprene excretion in men. *Lipids*, 1993;28: 705–8
33. Bajtarevic A, Ager C, Pienz M et al: Noninvasive detection of lung cancer by analysis of exhaled breath. *BMC Cancer*, 2009; 9: 348
34. Poli D, Carbognani P, Corradi M et al: Exhaled volatile organic compounds in patients with non-small cell lung cancer: Cross sectional and nested short-term follow-up study. *Respir Res*, 2005; 6: 71
35. Mazzone PJ, Hammel J, Dweik R et al: Diagnosis of lung cancer by the analysis of exhaled breath with a colorimetric sensor array. *Thorax*, 2007; 62: 565–68
36. Kischkel S, Miekisch W, Sawacki A et al: Breath biomarkers for lung cancer detection and assessment of smoking related effects – confounding variables, influence of normalization and statistical algorithms. *Clin Chim Acta*, 2010; 411: 1637–44
37. Greiter MB, Keck L, Siegmund T et al: Differences in exhaled gas profiles between patients with type 2 diabetes and healthy controls. *Diabetes Technol Ther*, 2010; 12: 455–63
38. Musaveloso K, Likhodii SS, Cunnane SC: Breath acetone is a reliable indicator of ketosis in adults consuming ketogenic meals. *Am J Clin Nutr*, 2002; 76: 65–70
39. Amal H, Ding L, Liu B et al: The scent fingerprint of hepatocarcinoma: *In-vitro* metastasis prediction with volatile organic compounds (VOCs). *Int J Nanomedicine*, 2012; 7: 4135–46