



Analysis of volatile organic compounds in exhaled breath for lung cancer diagnosis using a sensor system

DOI:

[10.1016/j.snb.2017.08.057](https://doi.org/10.1016/j.snb.2017.08.057)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Chang, J-E., Lee, D-S., Ban, S-W., Oh, J., Jung, M. Y., Kim, S-H., ... Jheon, S. (2017). Analysis of volatile organic compounds in exhaled breath for lung cancer diagnosis using a sensor system. *Sensors and Actuators B: Chemical: international journal devoted to research and development of physical and chemical transducers*. <https://doi.org/10.1016/j.snb.2017.08.057>

Published in:

Sensors and Actuators B: Chemical: international journal devoted to research and development of physical and chemical transducers

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Accepted Manuscript

Title: Analysis of volatile organic compounds in exhaled breath for lung cancer diagnosis using a sensor system

Authors: Ji-Eun Chang, Dae-Sik Lee, Sang-Woo Ban, Jaeho Oh, Moon Youn Jung, Seung-Hwan Kim, SungJoon Park, Krishna Persaud, Sanghoon Jheon



PII: S0925-4005(17)31479-X
DOI: <http://dx.doi.org/doi:10.1016/j.snb.2017.08.057>
Reference: SNB 22924

To appear in: *Sensors and Actuators B*

Received date: 15-10-2016
Revised date: 3-7-2017
Accepted date: 7-8-2017

Please cite this article as: Ji-Eun Chang, Dae-Sik Lee, Sang-Woo Ban, Jaeho Oh, Moon Youn Jung, Seung-Hwan Kim, SungJoon Park, Krishna Persaud, Sanghoon Jheon, Analysis of volatile organic compounds in exhaled breath for lung cancer diagnosis using a sensor system, *Sensors and Actuators B: Chemical* <http://dx.doi.org/10.1016/j.snb.2017.08.057>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Analysis of volatile organic compounds in exhaled breath for lung cancer diagnosis using a sensor system

Ji-Eun Chang^{1†}, Dae-Sik Lee^{3†}, Sang-Woo Ban⁴, Jaeho Oh⁴, Moon Youn Jung³, Seung-Hwan Kim³, SungJoon Park^{1§}, Krishna Persaud⁵, and Sanghoon Jheon^{1,2*}

¹ Department of Thoracic and Cardiovascular Surgery, Seoul National University Bundang Hospital, Seongnam-si, Gyeonggi-do, Republic of Korea

² Department of Thoracic and Cardiovascular Surgery, Seoul National University College of Medicine, Seoul, Republic of Korea

³ Bio-Medical IT Convergence Research Division, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

⁴ Department of Information and Communication Engineering, Dongguk University, Gyeongju, Gyeongbuk, Republic of Korea

⁵ School of Chemical Engineering and Analytical Science, The University of Manchester, Oxford Road, Manchester, United Kingdom

[†] These two authors contributed equally to this work.

[§] Current address: Department of Thoracic and Cardiovascular Surgery, Sanggye Paik Hospital, Inje University, Republic of Korea

* Correspondence to: Sanghoon Jheon, MD, PhD

Department of Thoracic and Cardiovascular Surgery, Seoul National University Bundang Hospital, 82,
Gumi-ro 173 beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, 463-707, Republic of Korea.

Tel: +82 31 787 2100; Fax: +82 31 787 2109; E-mail: jheon@snu.ac.kr

Highlights

- We have designed and fabricated a sensor system for early lung cancer diagnosis.
- Breath gas was collected from lung cancer patients before and after the surgery.
- Volatile organic compounds in the breath gas were analyzed by the sensor system.
- Lung cancer patients were distinguished from healthy controls with 75.0% accuracy.
- Breath analysis may develop as a novel diagnostic tool for early lung cancer.

ABSTRACT

Lung cancer is the leading cause of cancer deaths in worldwide. There are many challenges to detect early stage lung cancer in accurate, inexpensive, and non-invasive ways. In this study, we have designed, fabricated, and characterized a sensor system as a novel lung cancer diagnosis tool. In order to investigate the clinical feasibility of the system to detect early stage lung cancer, exhaled breath was collected from 37 patients with non-small cell lung cancer (81.1% of stage I and II) and 48 healthy controls. Three types of samples were collected from each patient; 1) before lung cancer surgery, 2) the first outpatient clinic visit after surgery, and 3) the second outpatient clinic visit after surgery. The sampling schedule from the healthy controls matched that of the lung cancer patients. The volatile organic compounds (VOCs) in the exhaled breath were analyzed by the sensor system. The sensor system consisted of an array of seven metal oxide gas sensors, a gas flow controlling module, heating module, gas adsorption-desorption module and classifiers for data analysis. The result obtained for the first set of samples, using a multilayer perceptron (MLP) for classification indicated a total accuracy of 75.0% with 79.0% of sensitivity and 72.0% of specificity. 93.5% of healthy controls showed nearly unchanged data from the first to the third samples, while 45.2% of lung cancer patients showed definitely changed data from the first to the third samples when analyzed

the projection results of original data onto the selected principal components (PCs) obtained from ~~by~~ principal component analysis (PCA). The study showed that VOCs in exhaled breath potentially discriminated mostly early stage lung cancer patients before the surgery from healthy volunteers with our sensor system. Furthermore, the prognosis of the lung cancer patients after surgery would be predicted by this system. These results suggest that breath analysis may develop as a novel diagnostic tool for early lung cancer.

Keywords: Lung cancer, Breath gas, Volatile organic compounds, Chemo-resistive sensors

1. Introduction

Lung cancer is one of the leading causes of death worldwide in both men and women [1]. Most lung cancer patients are diagnosed at an advanced stage when the symptoms (i.e., cough, dyspnea, fatigue, pain in thorax) [2] appear and this often leads to poor prognosis [3]. Pulmonary surgical resection is the most effective treatment for lung cancer, however, the 5-year survival rate after surgical resection of stage III patients is only 30% while stage I patients show up to 70% [4]. Therefore, earlier detection of lung cancer is the key to overcoming the disease. Various diagnosis tools for lung cancer such as chest x-ray, chest computed tomography (CT) scan, fluorodeoxyglucose-positron emission tomography (FDG-PET), bronchoscopy, and lung biopsy are applied. However, the existing diagnostic procedures are invasive, expensive, or inaccurate. Currently, low-dose CT screening is adopted but overdiagnosis often happens [5]. To improve disadvantages of the existing tools, there are many attempts to evaluate novel diagnostic tools for lung cancer. Detecting biomarkers from the blood [6, 7], sputum [8, 9], and bronchoalveolar lavage (BAL) fluid [10] were evaluated.

Breath gas is also known as a novel biomarker for lung cancer patients. In 1971, Pauling et al. first described breath gas analysis from a human being [11]. It is well known that breath gas contains various volatile organic compounds (VOCs), furthermore, lung cancer-related changes of VOCs are documented. Exhaled VOCs can originate from two main sources; exogenous volatiles that are inhaled and then exhaled and those endogenously produced by different biochemical processes through basic cellular functions [12]. It was suggested that activated cytochrome P450 (CYP) mixed oxidase enzymes in lung cancer patients may accelerate the degradation of several VOCs which are markers of oxidative stress and lead to alterations of VOCs in the breath during carcinogenesis [13-15]. The origin of exhaled VOCs is assumed to be mainly alveolar, however, direct comparison of VOC profiles from different parts of the lung and the airways is still missing.

Most of the studies reported analyzed VOCs using gas chromatography mass spectrometry (GC-MS) [16-19]. Though GC-MS is a powerful system which enables analysis of VOCs qualitatively and quantitatively, it is of high cost, requires skilled operators, and a lot of time [12, 20]. In addition, since

many different VOCs are detected from human exhaled breath by GC-MS, biomarkers from lung cancer patients are hard to define. To overcome the limitations of GC-MS analysis, various types of electronic nose have come into use for analyzing breath gas. In the previous studies, quartz microbalance (QMB) gas sensors [21], surface acoustic wave (SAW) gas sensors [22], colorimetric sensor array [23], polymer/carbon composite [24], conducting polymer gas sensors [25], ion mobility spectrometry [26], exhaled breath condensate (EBC) [27], and metal oxide gas sensors [28] were applied to lung cancer detection through breath gas. Among them, the metal oxide sensors are known to possess several advantages with high sensitivity, very fast response, and low cost [29]. The gas-flow control system with a sample pre-concentration module and signal processing module for pattern classification are also very essential components for diagnosing lung cancer by exhaled breath [26, 27].

In this study, we developed a system for analyzing VOCs from exhaled breath to detect early stage lung cancer, consisted of seven metal oxide gas sensors, gas condensation module with porous polymer sorption fiber, mass flow controller (MFC)-based gas flow controlling module using heaters and nitrogen carrier gas, and pattern analyzers. It was applied to 85 human breath samples (37 non-small cell lung cancer patients and 48 healthy controls) to distinguish following categories; lung cancer versus healthy control, before versus after surgery, sex, age, smoking history, histology, stage, tumor size, and tumor location.

2. Methods

2.1. Study population

The study was approved by the Institutional Review Board of Seoul National University Bundang Hospital (E-1208/167-004) and signed consent was obtained from the 85 volunteers (37 lung cancer patients, and 48 healthy controls) prior to the study. The patients who were diagnosed as lung cancer and underwent lung cancer surgery between July 2014 and December 2014 at Seoul National University Bundang Hospital were recruited. The patients with any other pulmonary disease than lung cancer and the healthy controls with any pulmonary disease were excluded from the study group. Sex, age, smoking history data for both groups and histology, pathologic staging by the tumor, node, metastasis (TNM) system according to the 7th edition of American Joint Committee on Cancer (AJCC) manual, tumor size and tumor location data for lung cancer group were collected based on questionnaire responses and medical records like CT images.

2.2. Study design

This study was a cross-sectional case-control study. For each subject, the exhaled breath was sampled thrice. The samples from the lung cancer patients were collected 1) before the lung cancer surgery (the operation day), 2) the first outpatient clinic visit after discharge (average 19.1 days after the first sampling), and 3) the second outpatient clinic visit after discharge (average 34.4 days after the second sampling). The sampling schedule from the healthy controls matched that of the lung cancer patients. For the lung cancer group, 37, 36, and 31 subjects were participated in only the first, the first to the second, and the first to the third sampling, respectively. For the healthy control group, 48, 46, and 46 subjects were participated in only the first, the first to the second, and the first to the third sampling, respectively. Some subjects were excluded during the study due to the cancelation of outpatient clinic or giving up participation of the study.

2.3. Exhaled breath collection and sampling

Eating, drinking, smoking, tooth-brushing with toothpaste or gargling with gargle were not allowed for minimum of 2 hour before sampling. The subjects deeply breathed into 3L Tedlar gas bag (Dongbang Hitech Company, Seoul, Korea) through a valve. To collect the alveolar breath and remove the dead space air, the first collected breath was discarded. Two valves were connected to the Tedlar bag; the one was used for the mouthpiece and the other one was applied to eliminate the dead volume. The desorption tube (Tenax TA, Gerstel, Baltimore, MD, USA) was flushed with 99.99% pure nitrogen gas which had been flushed three times before sampling. After flushing, the tube was connected with valve from the Tedlar bag at room temperature for 30 minutes in order to adsorb the VOCs in breath gas. During the adsorption process, we tapped the surface of the Tedlar bag constantly to adsorb the VOCs equally. The analysis was performed immediately after sampling from the subject.

2.4. Design and fabrication of sensor system

The sensor system consisted of seven metal oxide gas sensors (2600, 2602, 2610, 2611, 2620, 2442, 2444, Figaro Engineering Inc., Osaka, Japan) in an array, gas flow controlling module with three MFCs, two- or three-way valves, and a nitrogen carrier gas, heating module, gas adsorption-desorption module and classifiers (linear discriminant analysis; LDA, support vector machine; SVM, and multilayer perceptron; MLP) for data analysis. We chose the above commercialized sensors in our sensor system due to the several advantages such as long term stability, signal reliability, very fast response, and low cost.

2.5. Data analysis

The multi-dimensional sensing data measured by the sensor array was applied to pattern classification models for discriminating the lung cancer data from the control data. In the course of finding an

optimal classification model for this application, general well-known classification models were considered such as MLP, two-class SVM, one-class SVM and LDA [30, 31]. MLP is a popular supervised learning model for pattern classification. SVM is also a supervised learning pattern classification model that provides maximum marginal distance between two classes. LDA is an approach to look for an optimal hyper plane in terms of maximizing between class variance and minimizing within class variance. These general classification models might successfully discriminate the target class data (lung cancer patient data) from non-target class data (healthy control data).

Moreover, principal component analysis (PCA) was applied for analyzing distribution of high-dimensional sensing data, in which it was possible to examine the characteristics of discrimination between lung cancer data and healthy control data by a dimension reduction through projection onto the chosen principal components obtained from PCA. PCA seeks a projection that best represents the data in a least-squares sense, in which PCA projects d -dimension data onto a lower-dimension subspace in a way that is optimal in a sum-squared error sense [32]. First, the d -dimension mean vector μ and $d \times d$ covariance matrix Σ are computed for the full training data set of sensor array signals. Next, the eigenvectors \mathbf{e}_i and eigenvalues λ_i are computed, and sorted according to decreasing eigenvalue.

$$\lambda_1 > \lambda_2 > \dots > \lambda_d$$

The k eigenvectors having the largest eigenvalues are chosen as mainly contributive principal components. By projecting each d -dimension data onto the chosen k eigenvectors, k -dimension representation is generated for each d -dimension data. Therefore, each sensed higher-dimension signal is represented as a lower dimensional data by projection onto principal components.

The sensitivity and specificity were calculated using following equations. (TP; true positive, FP; false positive, TN; true negative, FN; false negative)

$$\text{Sensitivity (\%)} = \left(\frac{\text{TP}}{\text{TP} + \text{FN}} \right) \times 100$$

$$\text{Specificity (\%)} = \left(\frac{\text{TN}}{\text{TN} + \text{FP}} \right) \times 100$$

3. Results

3.1. Sensor characteristics

Fig. 1 shows a schematic diagram of the system for analyze exhaled breath gas (a), the photographs of the system (b), a desorption tubes filled with a porous material based on 2,6-diphenylene oxide polymer (Tenax TA) for pre-concentration of VOCs in the breath gas, removal of humidity and desorption of VOCs to the sensors (c), and a module for delivering the condensed samples to the test chamber with sensors (d). In the system, to acquire reproducible and high sensitively detected VOCs, the chamber temperature was designed to be controllable from room temperature to 300°C.

The response of the sensor array to the exhaled breath gas was directly related to the gas detection and pattern identification. First, a simulated lung cancer breath, which is a mixture of representative VOCs at concentrations similar to those determined by GC-MS analysis of lung cancer patient's exhaled breath, has been synthesized (a mixture of 145 ppb ethylbenzene, 24 ppb undecane, 67 ppb 4-methyl-octane, and 20 ppb 2,3,4-trimethyl-hexane with 80% relative humidity, 16% O₂, 5% CO₂ and 1.0 ppm CO) according to a previous report [23] and applied to the sensor array at 100 sccm rate to observe the response. As shown in Fig. 2, the gas reaction was completed after 260 seconds, whereas the desorption of the gas was completed after around 20 seconds. The sensors showed differentiated sensitivity curves to the VOCs according to the type. Since very low concentrations of VOC were presented in the real human breath, the responses of two sensors are relatively higher than other sensors which led to show only two sensors were working. The two sensors which showed higher sensitivities were TGS2600 and TGS2602, indicating VOCs such as iso-butane, toluene, ethanol were highly expressed in the real human breath samples. Though the other sensors showed relatively low signals, there were differences between lung cancer patients and healthy controls. They showed negative responses for lung cancer patients (Fig. 2b) while they showed positive responses for healthy controls (Fig. 2c). The response time of real human breath sample was longer than the simulated one because it takes some time to desorb gases from the porous polymer desorption fiber by flowing

nitrogen gas. Thus, the gas sensing properties were measured using a custom-made computer-controlled characterization system 10 minutes after the VOCs exposure. The gas response is defined as $R=(R_a-R_g)/R_a$, where R_a and R_g are the resistance in air and in a VOC atmosphere, respectively.

3.2. Subject characteristics

The characteristics of the lung cancer patients and the healthy controls are described in Table 1. Of the 37 non-small cell lung cancer patients, 27 were adenocarcinoma (73.0%), 9 squamous cell carcinoma (24.3%), and 1 basaloid carcinoma (2.7%). Among them, 25 were stage I (67.6%), 6 in stage II (16.2%), 5 in stage III (13.5%), and 1 in stage IV (2.7%). The smoking history was categorized as never smokers (< 100 lifetime cigarettes), former smokers (quit \geq 1 year before first sampling), or current smokers (quit < 1 year before first sampling). Though the lung cancer patients were imposed to quit smoking minimum 1 week before their lung cancer surgeries and maintain smoking cessation after the inner area surgery, their smoking history categories remained unchanged during the study. The tumor locations were divided into central and peripheral. When inner area more than two-thirds of the tumor diameter had contacted with the lung hilus, it was defined as the central tumor. 9 of total 37 lung cancer patients had central tumors (24.3%).

3.3. Analysis of VOCs in exhaled breath

The first samples measured from patients before surgery and healthy subjects were analyzed by projecting those onto principal components (PCs) obtained from PCA, in which the possibility of discrimination or distribution property of the samples of two classes will be verified by plotting 7-dimension data points onto the lower dimension (2-dimension) vector space constructed by the first PC and the second PC. As shown in Fig. 3, most of the projected data were highly correlated and it was very difficult to distinguish clearly between the groups for every category that we considered.

In order to compare the time varying characteristics of patient samples after surgery as time goes, PCA was separately conducted using each sample such as the first (before surgery), the second (the first outpatient clinic visit after surgery), and the third (the second outpatient clinic visit after surgery) samples together with healthy controlled subject samples. Then all samples of lung cancer patients and healthy controlled subjects were plotted onto the 2-dimension space constructed by the first PC and the second PC obtained from each corresponding PCA using the first samples, the second samples and the third samples, respectively. Fig. 4 shows that the lung cancer data had shifted to healthy control data zone as time goes after surgery.

When we analyzed every individual data, 93.5% (43/46) of healthy controls showed nearly unchanged data from the first, the second, and the third samples, while 45.2% (14/31) of lung cancer patients showed definite changed data in the first, the second, and the third samples. 14 of the lung cancer patients who showed shifted VOC patterns consisted of 4 stage IA (33.3%, 4 out of 12 total stage IA patients participated from the first to the third samplings), 6 stage IB (66.7%, 6 out of 9 total stage IB patients participated from the first to the third samplings), 2 stage IIA (40.0%, 2 out of 5 total stage IIA patients participated from the first to the third samplings), 1 stage IIIA (33.3%, 1 out of 3 total stage IIIA patients participated from the first to the third samplings), and 1 stage IV (100.0%, 1 out of 1 total stage IV patients participated from the first to the third samplings). The representative individual data from the first, the second, and the third samples are shown in Fig. 5.

To distinguish lung cancer patients from the healthy controls more clearly, two class-SVM and MLP analysis were performed with the first samples. The application of a two class-SVM found out 57.5% of total accuracy with 63.2% of sensitivity and 52.4% of specificity. By MLP analysis, the total accuracy was 75.0% with 79.0% of sensitivity and 72.0% of specificity. 1. In these experiments, a cross-validation was conducted for verifying the performance of the classifiers, in which a half each of the samples are selected as training data sets and the other half of them are selected as a test data set. Every accuracy figure shown in these experiments is a correct classification performance for a test data set of the classifier.

As MLP analysis showed higher total accuracy of distinguishing lung cancer patients from healthy controls with the first samples, we also analyzed the second and the third samples by MLP (Table 2). The sensitivity was decreased from the first to the third samples. The sensitivities of the first, the second, and the third samples were 79.0%, 60.0%, and 20.0%, respectively. The total accuracy was also decreased from the first to the third samples indicating 75.0%, 71.4%, and 61.2%, respectively.

We analyzed additional categories, i) lung cancer versus healthy control (by sex), ii) sex, iii) smoking history, iv) stage, and v) tumor location by MLP. The accuracies of each category are shown in Table 3. The accuracy of lung cancer versus healthy control was higher in males (74.1%) compared with females (56.3%). The accuracy of never and former smoker versus current smoker from total subjects and from lung cancer patients only were 81.4% and 79.0%, respectively. Stage I was distinguished from the other stages with 70.0% accuracy. The tumor location discrimination led to 61.1% accuracy from central to peripheral region.

4. Discussion

Electronic noses have been applied to various analysis fields such as food [33, 34] and various diseases; diabetes [35], uremia [36], asthma [37], chronic obstructive pulmonary disease [38], pulmonary sarcoidosis [39], colorectal cancer [40], and malignant mesothelioma [41, 42]. The application of electronic noses to detect lung cancer was also proposed. It is based on the lung cancer-related changes of VOCs in the exhaled breath gas from the lung cancer patients. The evidence of these alteration are suggested by metabolic and biochemical pathways [43, 44].

Many studies using GC-MS to analyze breath VOC biomarkers from lung cancer patients have been performed. Various VOCs were reported as lung cancer biomarkers such as acetone, benzene, heptane, hexanal, pentanal, pentane, styrene, and toluene [17, 45, 46]. However, accurate and reproducible lung cancer related breath VOCs has not been defined yet. Some VOCs found in lung cancer patients were known to be present in the environment and making it difficult to select true lung cancer biomarkers. Thus, it is very important to eliminate the interference of environmental VOCs to elevate the accuracy and reliability of the data.

In the present study, we adopted seven types of Taguchi Gas Sensors (TGS) which are known to possess low cost, fast response, and long term stability. Since very low concentrations of VOC biomarkers are present in the exhaled breath from the lung cancer patients, developing more sensitive materials was desirable. Comparing elevating the chamber temperature to the sensors' working temperatures of 300°C, using the pure nitrogen (99.99%) as a carrier gas was favorable to reliable detection by reducing humidity and environmental gas interferences. In addition, utilization of the porous material based on 2,6-diphenylene oxide polymer was also useful for the screening system of the lung cancer by pre-concentration of VOCs in the exhaled breath with removing humidity.

The breath gas sampling was one of the most critical factors in exhaled breath analysis. We standardized the sampling protocol to prepare the samples in the correct and consistent format. To eliminate any VOCs produced during digestion process, we asked the subjects to fast minimum of 2

hour before sampling. The first sampling from lung cancer patients, made just before the operation, therefore NPO (nothing by mouth) after midnight was performed. Smoking, applying toothpaste or mouthwash were also prohibited to minimize any chemical noise. Every sample was achieved in the same empty room to unify the background air. Subjects were encouraged to take a maximum possible deep inspiration followed by take a maximum possible deep expiration. The Tedlar gas bag was selected as a breath sample bag since it was a gas sampling bag made of polyvinylfluoride (PVF) with good stability for VOCs. In addition, it was proved not to possess its own VOCs by pilot studies. The analysis room was also strictly restricted to any chemicals which can affect the result.

Firstly, in this study dimension reduction was conducted using PCs obtained from PCA in order to verify possibility of data discrimination between lung cancer patients and healthy controls. The dimension reduction showed that it was hard to separate lung cancer patients from healthy controls. However, it was possible to indicate the individual VOC pattern alterations before and after lung cancer surgery. The samples from lung cancer patients were made both before and after lung cancer operation to find out if there is any VOC change after surgery. Interestingly, almost half of the lung cancer patients showed changed VOC patterns after surgery when analyzed by dimension reduction results using PCs obtained from PCA (Fig. 4).

It was highly supported by MLP analysis with decreased sensitivity after surgery (Table 2). In other words, after lung cancer surgery, and as time went by, it was hard to distinguish lung cancer patients from healthy controls. We carefully speculate that the prognosis of the lung cancer patients can be predicted by time-varying alteration of the dimension reduction results of the lung cancer patients samples onto the selected PCs obtained from PCA.

It was remarkable that 81.1% (30/37) of enrolled patients in this study were stage I and II lung cancer. Since the objective of the study was to detect 'early stage' lung cancer by breath gas analysis, results attained from the study were highly meaningful. In the previous studies, Machado et al. reported 7.1% (1/14) of stage I and II [47] and Peng et al. enrolled only stage III and IV [24] lung cancer patients.

In Table 3, we investigated various categories as many as possible. Never and former smokers versus current smokers were distinguished with high accuracy (81.4% from total subjects, 79.0% from lung cancer patients). It may suggest that smoking is a high influence factor for VOC alteration in human breath.

This study has some limitations. First of all, the age groups of the lung cancer and healthy control groups were different. We endeavored to minimize the age gap between the groups, however, most of the lung cancer patients were older than 60 years which made it difficult to find age-matched healthy controls without any pulmonary disease. Since no clear discrimination was shown according to the age (Fig. 3), we can speculate age differences between the groups rarely influenced the results. Secondly, even though we made an effort to restrict the sampling room and analysis room strictly to minimize any chemical noise, VOC-free air would have been better used in this study. Lastly, there need to be more smokers in the control group as a sensor system can detect such differences [48].

5. Conclusions

In conclusion, we have designed and fabricated a sensor system for lung cancer diagnosis or screening instrument consisting of a gas condensation module with porous polymer sorption fiber, flow control unit using embedded heaters and pure nitrogen carrier gas, seven metal oxide gas sensors in array, and pattern recognition engine. Using this sensor system, we successfully discriminated lung cancer patients (consisted of 81.1% of stage I and II) from healthy controls with 75.0% total accuracy by MLP analysis. In addition, the prognosis of the lung cancer patients after surgery could be predicted by this system based on dimension reduction using PCs obtained from PCA. These results suggest that VOCs in the exhaled breath gas analysis may develop as a novel diagnostic tool for early lung cancer or a prognostic tool after lung cancer resection surgery.

For further study, the correlation between the VOC changes of the lung cancer patients after lung cancer surgery and their prognosis should be elucidated.

Acknowledgements

This study was supported by the SNUBH Research Fund (14-2014-031), the Korean Foundation for Cancer Research (CB-2011-02-01), the National Research Foundation of Korea under research projects (NRF-2017M3A9F1033056), and the basic research program (12RC1510) of ETRI.

The authors are indebted to J. Patrick Barron, Professor Emeritus, Tokyo Medical University and Adjunct Professor, Seoul National University Bundang Hospital for his pro bono editing of this manuscript.

References

- [1] J. Ferlay, H.R. Shin, F. Bray, D. Forman, C. Mathers, D.M. Parkin, Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008, *Int J Cancer*, 127(2010) 2893-917.
- [2] H. Koyi, G. Hillerdal, E. Branden, A prospective study of a total material of lung cancer from a county in Sweden 1997-1999: gender, symptoms, type, stage, and smoking habits, *Lung Cancer*, 36(2002) 9-14.
- [3] R.S. Herbst, J.V. Heymach, S.M. Lippman, Lung cancer, *The New England journal of medicine*, 359(2008) 1367-80.
- [4] L. Dominioni, A. Imperatori, F. Rovera, A. Ochetti, G. Torrigiotti, M. Paolucci, Stage I nonsmall cell lung carcinoma, *Cancer*, 89(2000) 2334-44.
- [5] E.F. Patz, Jr., P. Pinsky, C. Gatsonis, J.D. Sicks, B.S. Kramer, M.C. Tammemagi, et al., Overdiagnosis in low-dose computed tomography screening for lung cancer, *JAMA internal medicine*, 174(2014) 269-74.
- [6] G. Tarro, A. Perna, C. Esposito, Early diagnosis of lung cancer by detection of tumor liberated protein, *J Cell Physiol*, 203(2005) 1-5.
- [7] E.F. Patz, Jr., M.J. Campa, E.B. Gottlin, I. Kusmartseva, X.R. Guan, J.E. Herndon, 2nd, Panel of serum biomarkers for the diagnosis of lung cancer, *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25(2007) 5578-83.
- [8] W.A. Palmisano, K.K. Divine, G. Saccomanno, F.D. Gilliland, S.B. Baylin, J.G. Herman, et al., Predicting lung cancer by detecting aberrant promoter methylation in sputum, *Cancer Res*, 60(2000) 5954-8.
- [9] Y.C. Wang, H.S. Hsu, T.P. Chen, J.T. Chen, Molecular diagnostic markers for lung cancer in sputum and plasma, *Annals of the New York Academy of Sciences*, 1075(2006) 179-84.
- [10] S.A. Ahrendt, J.T. Chow, L.H. Xu, S.C. Yang, C.F. Eisenberger, M. Esteller, et al., Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer, *J Natl Cancer Inst*, 91(1999) 332-9.
- [11] L. Pauling, A.B. Robinson, R. Teranishi, P. Cary, Quantitative analysis of urine vapor and breath

- by gas-liquid partition chromatography, *Proc Natl Acad Sci U S A*, 68(1971) 2374-6.
- [12] I. Horvath, Z. Lazar, N. Gyulai, M. Kollai, G. Losonczy, Exhaled biomarkers in lung cancer, *Eur Respir J*, 34(2009) 261-75.
- [13] M. Phillips, R.N. Cataneo, A.R. Cummin, A.J. Gagliardi, K. Gleeson, J. Greenberg, et al., Detection of lung cancer with volatile markers in the breath, *Chest*, 123(2003) 2115-23.
- [14] Y. Adiguzel, H. Kulah, Breath sensors for lung cancer diagnosis, *Biosensors and Bioelectronics*, 65(2015) 121-38.
- [15] Y. Hamuro, K.S. Molnar, S.J. Coales, B. OuYang, A.K. Simoneellis, T.C. Pochapsky, Hydrogen–deuterium exchange mass spectrometry for investigation of backbone dynamics of oxidized and reduced cytochrome P450 cam, *Journal of inorganic biochemistry*, 102(2008) 364-70.
- [16] C. Wang, R. Dong, X. Wang, A. Lian, C. Chi, C. Ke, et al., Exhaled volatile organic compounds as lung cancer biomarkers during one-lung ventilation, *Scientific reports*, 4(2014) 7312.
- [17] D. Poli, P. Carbognani, M. Corradi, M. Goldoni, O. Acampa, B. Balbi, 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*, 6(2005) 71.
- [18] M. Ligor, T. Ligor, A. Bajtarevic, C. Ager, M. Pienz, M. Klieber, et al., Determination of volatile organic compounds in exhaled breath of patients with lung cancer using solid phase microextraction and gas chromatography mass spectrometry, *Clinical chemistry and laboratory medicine : CCLM / FESCC*, 47(2009) 550-60.
- [19] M. Phillips, K. Gleeson, J.M.B. Hughes, J. Greenberg, R.N. Cataneo, L. Baker, et al., Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study, *The Lancet*, 353(1999) 1930-3.
- [20] P.J. Mazzone, Exhaled breath volatile organic compound biomarkers in lung cancer, *J Breath Res*, 6(2012) 027106.
- [21] C. Di Natale, A. Macagnano, E. Martinelli, R. Paolesse, G. D'Arcangelo, C. Roscioni, et al., Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors, *Biosensors & bioelectronics*, 18(2003) 1209-18.

- [22] X. Chen, M. Cao, Y. Li, W. Hu, P. Wang, K. Ying, et al., A study of an electronic nose for detection of lung cancer based on a virtual SAW gas sensors array and imaging recognition method, *Measurement Science and Technology*, 16(2005) 1535.
- [23] P.J. Mazzone, J. Hammel, R. Dweik, J. Na, C. Czich, D. Laskowski, et al., Diagnosis of lung cancer by the analysis of exhaled breath with a colorimetric sensor array, *Thorax*, 62(2007) 565-8.
- [24] G. Peng, U. Tisch, O. Adams, M. Hakim, N. Shehada, Y.Y. Broza, et al., Diagnosing lung cancer in exhaled breath using gold nanoparticles, *Nature nanotechnology*, 4(2009) 669-73.
- [25] J. Hatfield, P. Neaves, P. Hicks, K. Persaud, P. Travers, Towards an integrated electronic nose using conducting polymer sensors, *Sensors and Actuators B: Chemical*, 18(1994) 221-8.
- [26] H. Handa, A. Usuba, S. Maddula, J.I. Baumbach, M. Mineshita, T. Miyazawa, Exhaled breath analysis for lung cancer detection using ion mobility spectrometry, *PloS one*, 9(2014) e114555.
- [27] A.G. Dent, T.G. Sutedja, P.V. Zimmerman, Exhaled breath analysis for lung cancer, *J Thorac Dis*, 5(2013) S540-S50.
- [28] A. Berna, Metal oxide sensors for electronic noses and their application to food analysis, *Sensors*, 10(2010) 3882-910.
- [29] M. Mori, Y. Itagaki, J. Iseda, Y. Sadaoka, T. Ueda, H. Mitsuhashi, et al., Influence of VOC structures on sensing property of SmFeO₃ semiconductive gas sensor, *Sensors and Actuators B: Chemical*, 202(2014) 873-7.
- [30] S. Haykin, *Neural Networks: A Comprehensive Foundation*: Macmillan College Publishing Company, New York, (1994).
- [31] R. Duda, P. Hart, D. Stork, *Pattern classification*. 2nd edn Wiley, New York, (2000) 153.
- [32] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern classification*. 2nd, Edition New York, (2001).
- [33] C. Di Natale, A. Macagnano, F. Davide, A. D'Amico, R. Paolesse, T. Boschi, et al., An electronic nose for food analysis, *Sensors and Actuators B: Chemical*, 44(1997) 521-6.
- [34] S. Balasubramanian, S. Panigrahi, C. Logue, C. Doetkott, M. Marchello, J. Sherwood, Independent component analysis-processed electronic nose data for predicting *Salmonella typhimurium* populations in contaminated beef, *Food Control*, 19(2008) 236-46.

- [35] P. Wang, Y. Tan, H. Xie, F. Shen, A novel method for diabetes diagnosis based on electronic nose, *Biosensors & bioelectronics*, 12(1997) 1031-6.
- [36] Y.-J. Lin, H.-R. Guo, Y.-H. Chang, M.-T. Kao, H.-H. Wang, R.-I. Hong, Application of the electronic nose for uremia diagnosis, *Sensors and Actuators B: Chemical*, 76(2001) 177-80.
- [37] S. Dragonieri, R. Schot, B.J. Mertens, S. Le Cessie, S.A. Gauw, A. Spanevello, et al., An electronic nose in the discrimination of patients with asthma and controls, *The Journal of allergy and clinical immunology*, 120(2007) 856-62.
- [38] A.D. Hattesoehl, R.A. Joerres, H. Dressel, S. Schmid, C. Vogelmeier, T. Greulich, et al., Discrimination between COPD patients with and without alpha 1-antitrypsin deficiency using an electronic nose, *Respirology*, 16(2011) 1258-64.
- [39] S. Dragonieri, P. Brinkman, E. Mouw, A.H. Zwinderman, P. Carratu, O. Resta, et al., An electronic nose discriminates exhaled breath of patients with untreated pulmonary sarcoidosis from controls, *Respiratory medicine*, 107(2013) 1073-8.
- [40] E. Westenbrink, R.P. Arasaradnam, N. O'Connell, C. Bailey, C. Nwokolo, K.D. Bardhan, et al., Development and application of a new electronic nose instrument for the detection of colorectal cancer, *Biosensors & bioelectronics*, 67(2015) 733-8.
- [41] S. Dragonieri, M.P. van der Schee, T. Massaro, N. Schiavulli, P. Brinkman, A. Pinca, et al., An electronic nose distinguishes exhaled breath of patients with Malignant Pleural Mesothelioma from controls, *Lung Cancer*, 75(2012) 326-31.
- [42] E.A. Chapman, P.S. Thomas, E. Stone, C. Lewis, D.H. Yates, A breath test for malignant mesothelioma using an electronic nose, *European Respiratory Journal*, 40(2012) 448-54.
- [43] P.J. Mazzone, X.-F. Wang, Y. Xu, T. Mekhail, M.C. Beukemann, J. Na, et al., Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer, *Journal of Thoracic Oncology*, 7(2012) 137-42.
- [44] S. Subramaniam, R.K. Thakur, V.K. Yadav, R. Nanda, S. Chowdhury, A. Agrawal, Lung cancer biomarkers: State of the art, *Journal of carcinogenesis*, 12(2013) 3.
- [45] S. Kischkel, W. Miekisch, A. Sawacki, E.M. Straker, P. Trefz, A. Amann, et al., Breath

biomarkers for lung cancer detection and assessment of smoking related effects—confounding variables, influence of normalization and statistical algorithms, *Clinica Chimica Acta*, 411(2010) 1637-44.

[46] G. Peng, M. Hakim, Y.Y. Broza, S. Billan, R. Abdah-Bortnyak, A. Kuten, et al., Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors, *Br J Cancer*, 103(2010) 542-51.

[47] R.F. Machado, D. Laskowski, O. Deffenderfer, T. Burch, S. Zheng, P.J. Mazzone, et al., Detection of lung cancer by sensor array analyses of exhaled breath, *Am J Respir Crit Care Med*, 171(2005) 1286-91.

[48] Z. Cheng, G. Warwick, D. Yates, P. Thomas, An electronic nose in the discrimination of breath from smokers and non-smokers: a model for toxin exposure, *Journal of breath research*, 3(2009) 036003.

Biographies

Ji-Eun Chang, PhD is a senior researcher in the Department of Thoracic and Cardiovascular Surgery at Seoul National University Bundang Hospital. She is a pharmacist and her research interests include lung cancer biomarkers.

Dae-Sik Lee, PhD is a principal researcher in Electronics and Telecommunications Research Institute. He is a professor in University of Science and Technology.

Sang-Woo Ban, PhD is a professor in the Department of Information and Communication Engineering at Dongguk University.

Jaeho Oh is a PhD candidate in the Department of Information and Communication Engineering at Dongguk University.

Moon Youn Jung, PhD is a principal researcher in Electronics and Telecommunications Research Institute.

Seung-Hwan Kim, PhD is a principal researcher in Electronics and Telecommunications Research Institute.

SungJoon Park, MD is a professor in the Department of Thoracic and Cardiovascular Surgery at Sanggye Paik Hospital, Inje University. He is a general thoracic surgeon.

Krishna Persaud, PhD is a professor in the School of Chemical Engineering and Analytical Science at the University of Manchester.

Sanghoon Jheon, MD, PhD is a President of Seoul National University Bundang Hospital. He is a professor in the Department of Thoracic and Cardiovascular Surgery at Seoul National University College of Medicine. He is a general thoracic surgeon.

Figure legends

Figure 1. Schematic diagram of the exhaled breath gas measuring system (a), the photographs of the completed system (b), desorption tubes filled with a porous material based on 2,6-diphenylene oxide polymer (c), and a delivery module (d).

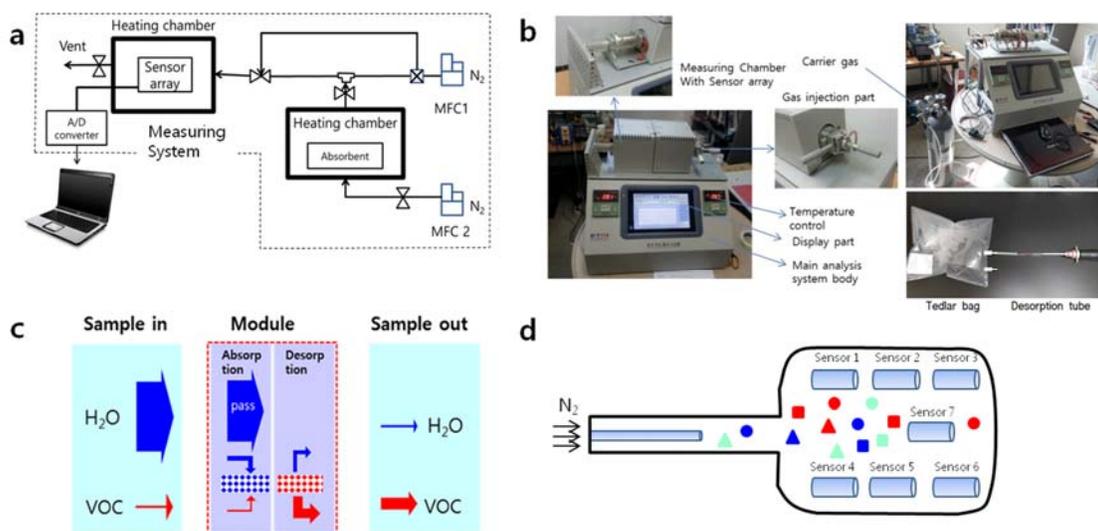


Figure 2. Response of the sensors to the simulated VOCs (a), VOCs in the breath gas from a lung cancer patient (b), and VOCs in the breath gas from a healthy control (c).

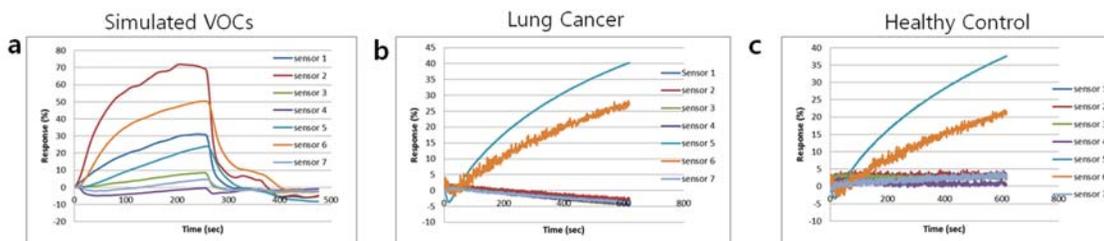


Figure 3. Dimension reduction using PCs obtained from PCA of the first samples categorized by i) lung cancer versus healthy control, ii) sex, iii) age, iv) smoking history, v) histology, vi) stage, vii) tumor size, and viii) tumor location.

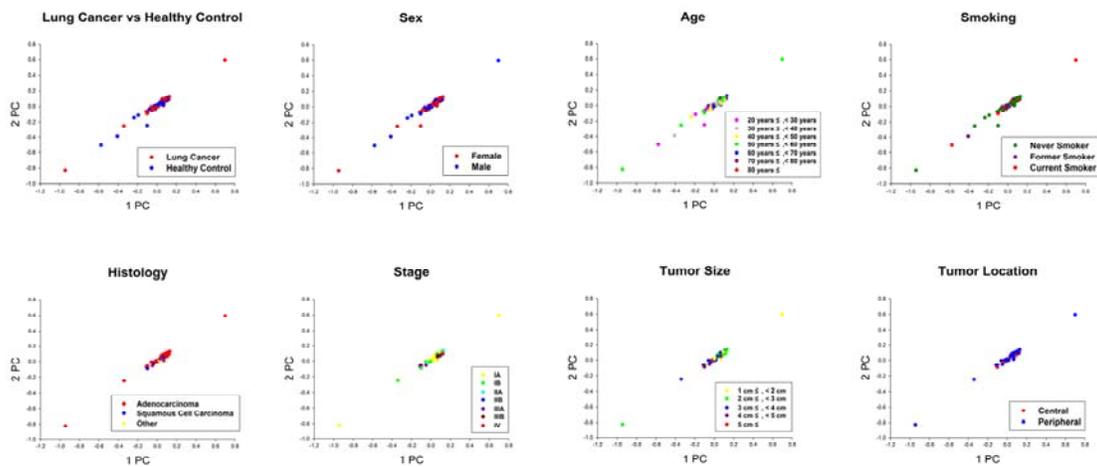


Figure 4. Dimension reduction using PCs obtained from PCA of the first (before surgery), the second (the first outpatient clinic visit after surgery), and the third (the second outpatient clinic visit after surgery) samples.

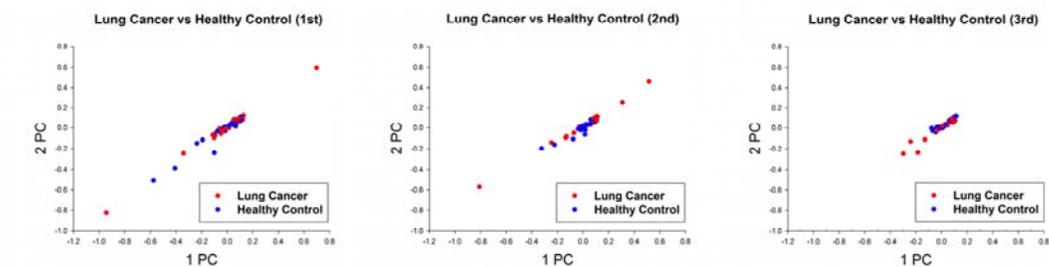


Figure 5. The representative individual data from the first, the second, and the third samples by dimension reduction using PCs obtained from PCA. (LC; Lung Cancer, H; Healthy Control)

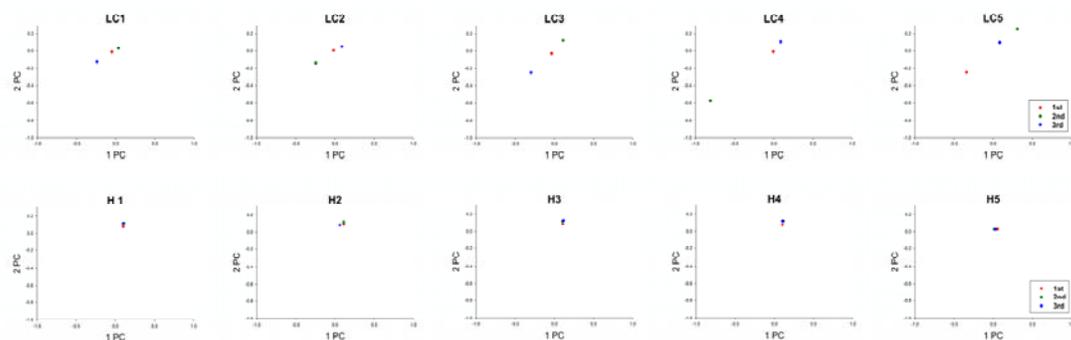


Table 1. The characteristics of the lung cancer patients and the healthy controls.

	Lung Cancer (<i>n</i> = 37)		Healthy Control (<i>n</i> = 48)	
Sex	Male	25 (67.6%)	Male	28 (58.3%)
	Female	12 (32.4%)	Female	20 (41.7%)
Age (years)	63.3 ± 10.4		38.7 ± 11.2	
Smoking history	Never	13 (35.1%)	Never	34 (70.8%)
	Former	15 (40.5%)	Former	10 (20.8%)
	Current	9 (24.3%)	Current	4 (8.3%)
Histology	Adenocarcinoma	27 (73.0%)		
	Squamous	9 (24.3%)		
	Other	1 (2.7%)		
Stage	IA	15 (40.5%)		
	IB	10 (27.0%)		
	IIA	5 (13.5%)		
	IIB	1 (2.7%)		
	IIIA	4 (10.8%)		
	IIIB	1 (2.7%)		
	IV	1 (2.7%)		
Tumor size (cm)	3.0 ± 1.1			
Tumor location	Central	9 (24.3%)		
	Peripheral	28 (75.7%)		

Table 2. MLP analysis of the first (before surgery), the second (the first outpatient clinic visit after surgery), and the third (the second outpatient clinic visit after surgery) samples.

(LC; Lung Cancer, H; Healthy Control)

	1st sample (LC; $n = 37$, H; $n = 48$)	2nd sample (LC; $n = 36$, H; $n = 46$)	3rd sample (LC; $n = 31$, H; $n = 46$)
Sensitivity (%)	79.0	60.0	20.0
Specificity (%)	72.0	79.3	89.7
Accuracy (%)	75.0	71.4	61.2

Table 3. Accuracies of various categories analyzed by MLP.

(Total = Lung Cancer + Healthy Control)

Lung Cancer vs Healthy Control		Male vs Female		Smoking History								Stage	Tumor Location			
Male	Female	Lung Cancer	Healthy Control	Never vs Former, Current		Never, Former vs Current		Never vs Former		Never vs Current		Former vs Current		I vs II,III,IV	Central vs Peripheral	
				Total	Lung Cancer	Total	Lung Cancer	Total	Lung Cancer	Total	Lung Cancer	Total	Lung Cancer			
Accuracy (%)	74.1	56.3	52.6	41.7	48.8	63.2	81.4	79.0	63.2	50.0	65.7	66.7	78.6	63.6	70.0	61.1