Detection of Lung Cancer by Sensor Array Analyses of Exhaled Breath

Roberto F. Machado, Daniel Laskowski, Olivia Deffenderfer, Timothy Burch, Shuo Zheng, Peter J. Mazzone, Tarek Mekhail, Constance Jennings, James K. Stoller, Jacqueline Pyle, Jennifer Duncan, Raed A. Dweik, and Serpil C. Erzurum

Departments of Pathobiology and Pulmonary, Allergy, and Critical Care Medicine, Lerner Research Institute, and Department of Hematology and Medical Oncology, Cleveland Clinic Foundation, Cleveland, Ohio; and Smiths Detection, Inc., Pasadena, California

Rationale: **Electronic noses are successfully used in commercial applications, including detection and analysis of volatile organic compounds in the food industry.** *Objectives:* **We hypothesized that the electronic nose could identify and discriminate between lung diseases, especially bronchogenic carcinoma.** *Methods:* **In a discovery and training phase, exhaled breath of 14 individuals with bronchogenic carcinoma and 45 healthy control subjects or control subjects without cancer was analyzed. Principal components and canonic discriminant analysis of the sensor data was used to determine whether exhaled gases could discriminate between cancer and noncancer. Discrimination between classes was performed using Mahalanobis distance. Support vector machine analysis was used to create and apply a cancer prediction model prospectively in a separate group of 76 individuals, 14 with and 62 without cancer.** *Main Results:* **Principal components and canonic discriminant analysis demonstrated discrimination between samples from patients with lung cancer and those from other groups. In the validation study, the electronic nose had 71.4% sensitivity and 91.9% specificity for detecting lung cancer; positive and negative predictive values were 66.6 and 93.4%, respectively. In this population with a lung cancer prevalence of 18%, positive and negative predictive values were 66.6 and 94.5%, respectively.** *Conclusion:* **The exhaled breath of patients with lung cancer has distinct characteristics that can be identified with an electronic nose. The results provide feasibility to the concept of using the electronic nose for managing and detecting lung cancer.**

Keywords: breath tests; bronchogenic cancer; electronic nose; volatile organic compounds

Smelling to establish diagnoses is a time-honored practice in medicine. For example, detecting fetor hepaticus and the putrid smell of anaerobic infections represent but two examples of olfactory diagnosis, which has largely been abandoned in the face of new diagnostic technologies. However, recent advances in odor-sensing technology, signal processing, and diagnostic algorithms have created chemical sensing and identification devices called "electronic noses," which promise to resurrect olfaction as an important diagnostic option. Electronic noses rely on arrays of chemical vapor sensors that respond to specific stereochemical characteristics of an odorant molecule, particularly volatile organic compounds (VOCs) (1). Multidimensional data obtained from the sensor array can be analyzed by statistical algorithms (e.g., principal components analysis, discriminant function analysis, factor analysis) or by structural algorithms (neural networks) to discriminate and identify odorant samples (2–4). Like the human nose, its electronic counterpart responds in concert to a given odor to generate a pattern, or "smellprint," which is analyzed, compared with stored patterns, and recognized.

Human breath contains a mixture of hundreds of VOCs (5), which offers the possibility that this new electronic nose technology may have many medical applications (6–8). There may be potential utility for electronic nose technology in medical applications, including identification of bacterial pathogens (6, 7, 9, 10) and pneumonia (8), and monitoring of glucose control in patients with diabetes (11). In this context, many VOCs, in particular alkanes and benzene derivatives, measured by mass spectrometry of the exhaled breath have been used to predict the presence of lung cancer in patients (12, 13). However, the method of mass spectrometry to separate and identify 20 or more VOCs in a complex mixture is cumbersome and requires expensive equipment and highly skilled analysts, which limits its widespread application in screening and diagnosis (14).

Because the electronic nose is highly sensitive for detecting VOCs, and based on a previous study involving patients with lung neoplasms (15), we hypothesized that an electronic nose would detect lung cancer on the basis of the complex smellprints of numerous VOCs in exhaled breath from individuals with lung cancer as compared with individuals with other, noncancer lung diseases, or healthy control subjects. Here, we applied support vector machine (SVM) analysis of smellprints of exhaled gases to create a cancer prediction model using a training set of exhaled breath from individuals with cancer or other noncancer lung diseases, or healthy control subjects. To validate the potential utility of smellprint signatures for identifying lung cancer, the discrimination power of the model was tested in an independent sample of 76 individuals. Some of the results of these studies have been previously reported in the form of an abstract (16).

METHODS

Study Population

We used two independent sets of volunteers for our studies: one for the discovery/training set to create the cancer prediction model and a second group for validating the model. The study population included individuals with bronchogenic carcinoma, healthy control subjects, and well-characterized patients with lung parenchymal, airway, or pulmonary vascular disease from subspecialty pulmonary clinics at the Cleveland Clinic Foundation, including emphysema caused by α_1-antitrypsin (AT) deficiency, chronic granulomatous disease caused by beryllium exposure, asthma, pulmonary hypertension, and smoking-related chronic obstructive pulmonary disease (COPD). Diagnosis of lung cancer was based on histologic confirmation. All individuals with α_1-AT deficiency were homozygous for the Z α_1-AT allele (PI*ZZ phenotype) and had serum α_1-AT levels below the protective threshold value of 11 μmol/L. Diagnosis of chronic pulmonary beryllium disease (CBD) was based on a positive lymphocyte proliferation test for beryllium and the presence

(*Received in original form September 8, 2004; accepted in final form February 24, 2005*)

Supported by the National Institutes of Health grant NIH HL04265 and Smiths Detection, Inc., through donation of the electronic nose instrument and partial financial support for research supplies.

Correspondence and requests for reprints should be addressed to Serpil C. Erzurum, M.D., Cleveland Clinic Foundation, 9500 Euclid Avenue/NB40, Cleveland, OH 44195. E-mail: erzurus@ccf.org

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 171. pp 1286–1291, 2005
Originally Published in Press as DOI: 10.1164/rccm.200409-1184OC on March 4, 2005
Internet address: www.atsjournals.org

of nonnecrotizing granulomata in transbronchial biopsy specimens. Asthma was defined according to the National Asthma Education and Prevention Program Expert Panel Report II guidelines (17). Pulmonary hypertension was defined as a mean resting pulmonary arterial pressure of greater than 25 mm Hg by right heart catheterization and classified according to the World Health Organization classification of pulmonary hypertension (18). COPD was defined according to the Global Initiative for Chronic Obstructive Pulmonary Disease (19). Healthy control individuals were identified by absence of pulmonary symptoms, history of pulmonary disease, and normal lung function. Patients with any acute disease exacerbation, history of cancer, and any active medical condition, such as diabetes, immunosuppression, or coronary artery disease, were excluded from the study. The study was approved by Institutional Review Board of the Cleveland Clinic Foundation, and informed consent was obtained from volunteers.

Study Design

The study was designed in two phases. The discovery and training phase included an initial sample of patients with lung cancer, healthy volunteers, and volunteers with lung disease. In this first phase, a nonblinded analysis of the exhaled breath of individuals was performed to explore possible differences between the study groups. When it was determined that the individuals with lung cancer had a distinct exhaled breath profile, the training set was used to create a model to be validated in the second, validation, phase of the study in which patients with lung cancer, healthy control subjects, and control subjects with other lung disease were evaluated in a cross-sectional blinded manner. All individuals evaluated during the discovery phase were included in the prediction model used in the second phase of the study. In the model, patients with chronic nonneoplastic lung disease were combined with healthy volunteers as "noncancer" control subjects.

Collection of Exhaled Breath

After exhalation to residual volume, the subject inhaled to total lung capacity through a mouthpiece that contained a cartridge on the inspiratory port (Cartridge N7500-2; North Safety Products, Cranston, RI), which removed more than 99.99% of VOCs from the air during inspiration, thus clearing the inhaled air of any ambient contaminants. Individuals exhaled against 10 cm H_2O pressure to ensure closure of the vellum to exclude nasal entrainment of gas. The exhaled gas was collected through a separate exhalation port of the mouthpiece in a nonreactive Mylar gas-sampling bag (20). A minimum of five analyses was performed on the exhaled breath of each volunteer.

Sensor Characteristics

Exhaled breath samples were analyzed by a handheld electronic nose, containing 32 polymer composite sensors, a sampling system, a data acquisition system, and a processor (Cyranose 320, Smiths Detection, Pasadena, CA). The sampling system delivers ambient air and the sample vapor to the sensors in sequence. The processor and embedded pattern-recognition software collect and analyze the differential responses of the sensors to the sample vapor. Each sensor of the array undergoes a reversible change in electrical resistance when exposed to a vapor or analyte. Furthermore, resistance change of each sensor is unique because of chemical diversity of the sensor materials. Consequently, a pattern of resistance changes is obtained from the sensor array to a given vapor, which is termed a smellprint.

Data Processing and Analysis for Discovery and Training Phase

Sensor response data were processed using Savitzky-Golay filtering and baseline correction (21). Different processing methods, such as normalization and scaling, were applied to determine the best discrimination among samples. In the discovery phase, sensor response data were analyzed using principal components analysis to reduce the data from 32 individual responses to vectors or principal components (22). The vectors were calculated to capture the maximum amount of variance in the dataset. The results were plotted in two dimensions using the first two vectors calculated. Statistical analysis using standard measures, such as p values, is not applicable to this multivariate data analysis. Rather, results from principal components analysis were used in canonic discriminant analysis to create a model that maximizes the distance among the two sample classes (i.e., cancer versus noncancer) (23). Quantification of the discrimination between two sample classes was performed using the Mahalanobis distance, a measure of class separation between different data clusters (23). A Mahalanobis distance of 3 or greater indicates that the classes are discrete from each other, and that the sensor array may be able to classify unknown samples belonging to different groups.

SVM Analysis

SVM analysis is a learning algorithm that can perform binary classification, or pattern recognition, by nonlinearly mapping n-dimensional input space into a high-dimensional feature space. In this high-dimensional feature space, a linear classifier, or nonlinear kernel classifier, is constructed, and the model is used to discriminate samples belonging to two different groups. Thus, an SVM learns to discriminate between the members and the nonmembers of a class. After learning the features of the class, the SVM recognizes unknown samples as a member of a specific class. SVMs have been shown to perform especially well in multiple areas of biological analyses, especially functional class prediction from microarray gene expression data and chemometrics (24–28). We constructed an SVM classifier with a nonlinear algorithm with Matlab (version 6.5) (Mathworks, Natick, MA) using the training set of sensor response data from subjects with lung cancer, subjects with noncancer disease, and healthy control subjects. Model parameters were optimized using the training set for maximum margin of class separation and minimum training set error ($C = 10$, width of gaussian $\sigma = 5$; for review, *see* Reference 20). The model was applied to a group of unknown samples from a different group of patients with lung cancer, healthy control subjects, and control subjects with other lung disease (i.e., noncancer volunteers). Each participant's exhaled breath was analyzed five times, and the results of the five separate sensor response data considered as the final determinant of lung cancer. In 92% of cases, outcomes of the five samplings were concordant. In the few cases with discordant responses, cancer or noncancer was predicted on the basis of the predominant response for that particular patient (e.g., if one individual had three responses as cancer and two as noncancer, the final assignment was cancer). Sensitivity, specificity, and positive and negative predictive values of the model for the diagnosis of lung cancer were determined. Details of the electronic nose and glossary of terms are provided in an online supplement.

Gas Chromatography and Mass Spectroscopy

Exhaled gas samples from individuals with lung cancer were collected in Tedlar sample bags (SKC, Inc., Eighty Four, PA) and analyzed on a Finnigan Trace gas chromatograph coupled to a Polaris Q Ion Trap Mass Spectrometer (Thermo Electron Corp., Woburn, MA). The gas chromatograph/mass spectrometer was optimized to target molecules with 4 to 12 carbons. Between 500 and 1,000 cm^3 of breath was concentrated on solid-phase thermal desorption tubes and delivered to gas chromatograph/mass spectrometer using an Entech 7100 gas concentrator (Entech Instruments, Simi Valley, CA).

Clinical Data Analysis

Quantitative data are summarized as mean ± SE; categoric data are summarized by frequencies. Two-tailed *t*-test statistics, χ^2, analysis of variance, and analysis of variance on ranks were used where appropriate, with the Bonferroni correction being applied to the significance criterion once pairwise comparisons were made among the study groups.

RESULTS

Discovery and Training

Fourteen individuals with bronchogenic carcinoma (one with small cell carcinoma and 13 with non–small cell carcinoma), 19 with α_1-AT deficiency, six with CBD, and 20 healthy control subjects were included in the discovery phase (Table 1). Patients with lung cancer were older and had smoked longer than individuals with α_1-AT deficiency, CBD, and control subjects ($p < 0.001$). Patients with lung cancer had similar FEV_1 ($p > 0.05$)

TABLE 1. CLINICAL CHARACTERISTICS OF THE STUDY POPULATION FOR TRAINING SET

	Lung Cancer	Control	Chronic Beryllium Disease	α_1-AT Deficiency
No. patients	14	20	6	19
Age, yr*	64 ± 3	38 ± 2	53 ± 7	50 ± 2
Sex, M/F	10/4	6/14	4/2	8/11
Tobacco use				
Ever smoker	14	7	1	15
Current smoker	2	6	0	0
Pack-years*	64 ± 13	12 ± 4	20	19 ± 3
Histologic type				
Non–small cell	13			
Small cell	1			
Pathologic stage				
IB	1			
IIIA	3			
IIIB	3			
IV	7			
FEV_1, % predicted*	56 ± 5	99 ± 6	84 ± 7	52 ± 8
FVC, % predicted*	66 ± 6	98 ± 6	90 ± 9	83 ± 5

Definition of abbreviations: α_1-AT = α1-antitrypsin; M/F = male/female.
All values are mean ± SEM.
* $p < 0.05$ by analysis of variance.

and FVC ($p > 0.05$) to α_1-AT deficiency patients, but lower FEV_1 ($p = 0.007$) than individuals with CBD.

Principal components analysis was used as an exploratory technique to investigate clustering of datasets within the multisensor space. Figure 1 demonstrates discrimination between samples from lung cancer, healthy control, α_1-AT deficiency, and CBD groups. In contrast to patients with α_1-AT deficiency and CBD, patients with lung cancer clustered distinctly from control subjects. No changes in clustering of samples were observed when patients were grouped by histologic type, pathologic stage, or severity of lung dysfunction. No clustering differences were demonstrated when current smokers and nonsmoking individuals were compared for healthy or disease states, including lung cancer. Therefore, discrimination of subjects with lung cancer is likely related to the disease process and not to smoking.

Canonic discriminant analysis was performed on the data. With the optimal number of vectors, the cross-validation results were 71.6% correct, with a Mahalanobis distance of 3.25 between individuals with lung cancer and normal individuals, as opposed to a Mahalanobis distance of 0.96 for α_1-AT deficiency and 1.5 for patients with CBD. On the basis of these results, we hypothesized that analysis of exhaled breath by the electronic nose could identify cancer from noncancer.

Fourteen patients with cancer, 19 patients with α_1-AT deficiency, six patients with CBD, two patients with COPD, and 20 healthy control subjects were included in the training phase to create the SVM algorithm.

Model Validation

Healthy control subjects and volunteers with lung cancer, COPD, asthma, or pulmonary hypertension were prospectively enrolled in the blinded validation study of the cancer prediction model. Exhaled breath was collected from 14 patients with active, nonresected, untreated lung cancer (Table 2), and from control groups, including the following: 30 nonsmoking healthy volunteers (age, 37 ± 3 years; 11 women), 12 patients with COPD (age, 66 ± 2 years; eight women; FEV_1, 46 ± 8% predicted; FVC, 68 ± 6% predicted; FEV_1/FVC, 0.45 ± 0.08), two patients with resected lung cancer in remission (both non–small cell carcinoma, stages IA and IIA), 11 patients with asthma (age, 42 ± 4 years; eight women; FEV_1, 82 ± 6% predicted; FVC, 90 ± 6% predicted; FEV_1/FVC, 0.73 ± 0.03), and seven patients with pulmonary hypertension (age, 48 ± 7 years; four women; primary, three patients; scleroderma-related, one patient;, intracardiac shunt, one patient; chronic thromboembolic disease, one patient; sarcoid vasculitis, one patient; mean resting pulmonary arterial pressure, 40 ± 7 mm Hg). The model correctly identified 330 of 388 independently analyzed exhaled breath samples, with an overall accuracy of 85% (95% confidence interval [CI], 81.1–88.2%; Figure 2).

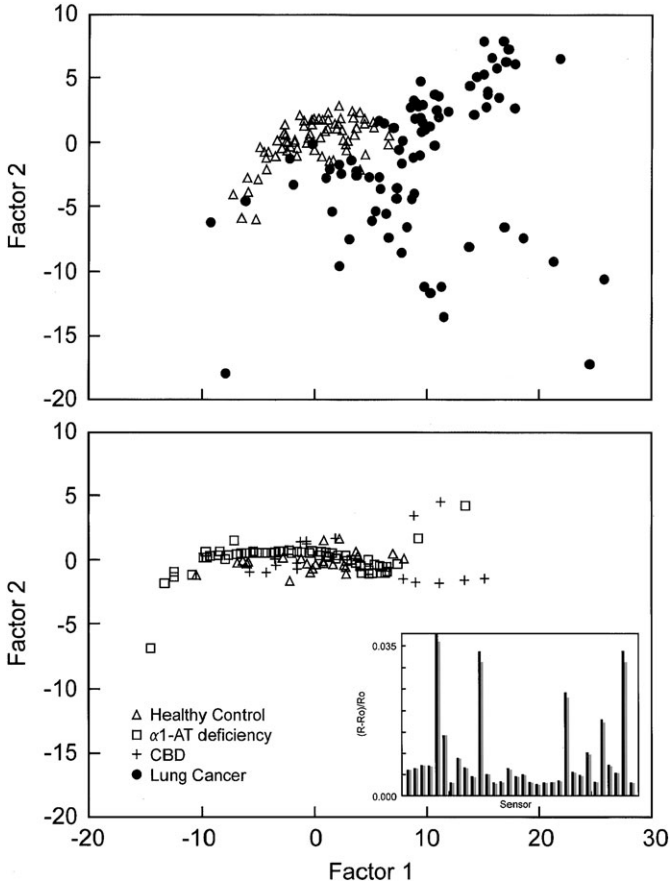

Figure 1. Principal components analysis plot, shown as a two-dimensional projection of the 32-dimensional vector analyses, demonstrates distinct clustering of the samples from patients with lung cancer separate from healthy control subjects (*upper panel*), whereas patients with interstitial lung disease or emphysema are not separable from healthy control subjects (*lower panel*). *Inset*: Example of a typical smellprint derived from the 32 sensor responses, which are used in the multidimensional analyses, from a healthy control subject (*black bars*) and a patient with lung cancer (*gray bars*). α1-AT = α_1-antitrypsin; CBD = chronic pulmonary beryllium disease; R = postmeasurement sensor resistance; Ro = baseline sensor resistance.

The electronic nose had 71.4% sensitivity and 91.9% specificity. In this population with an 18% prevalence of lung cancer, the positive predictive value of the algorithm for lung cancer was 66.6%, and negative predictive value for diagnosis of lung cancer was 93.4% (Table 3). Twenty-eight control patients with noncancer lung disease were correctly identified as noncancer (93.3%; 95% CI, 66.7–99.2). Two of the four false-negative identifications occurred in patients with small cell cancer. Because the original training set consisted primarily of patients with non–small cell lung cancer, this could suggest that our cancer prediction model is not optimal for evaluation of patients with suspected

TABLE 2. CLINICAL CHARACTERISTICS OF PATIENTS WITH LUNG CANCER IN THE CROSS-SECTIONAL VALIDATION STUDY

Patient No.	Age (*yr*)	Sex	Histologic Type*	Primary Lesion Size (*cm*)	Stage
1	75	Female	NSCCA	3	IB
2	57	Male	NSCCA	7	IIIB
3	42	Male	SCCA	1.7	Extensive
4	57	Female	NSCCA	2	IV
5	64	Male	SCCA	0.2	Extensive
6	63	Female	SCCA	2	Limited
7	55	Male	SCCA	1.9	Limited
8	65	Male	SCCA	3.5	Limited
9	56	Male	NSCCA	4.0	IV
10	65	Male	SCCA	3.5	Limited
11	61	Male	NSCCA	4.0	IB
12	85	Female	NSCCA	2.8	IA
13	76	Male	NSCCA	2	IIIA
14	34	Male	NSCCA	3	IIA

Definition of abbreviations: NSCCA = non–small cell lung cancer; SCCA = small cell lung cancer.

small cell cancer. The other false-negative results occurred in two patients who underwent diagnostic mediastinoscopy before exhaled breath assessment and had small primary and limited-stage lesions without airway involvement, suggesting that the model may not be ideally sensitive. False-positive results occurred in one patient with asthma with severe airflow limitation (FEV_1, 0.88, 33% of predicted; FVC, 1.73, 52% of predicted) and in one patient with primary pulmonary hypertension (resting mean pulmonary artery pressure, 50 mm Hg).

The results of the model were reproducible. Five healthy control subjects were repeatedly measured on two to six different occasions, and one patient with lung cancer had measurements done twice, and in all cases, class assignment was correct (data not shown).

Gas Chromatography and Mass Spectroscopy

To characterize VOCs in samples detected by the electronic nose, we performed gas chromatography and mass spectroscopy on exhaled breath of eight individuals with lung cancer from our prospective cohort. Table 4 lists selected VOCs measured and average concentration in parts per billion volume. Of VOCs present, many have been previously reported in different concentrations in patients with lung cancer compared with those without lung cancer (12–14).

DISCUSSION

The current study shows that exhaled breath of patients with lung cancer has distinct characteristics that can be identified with an

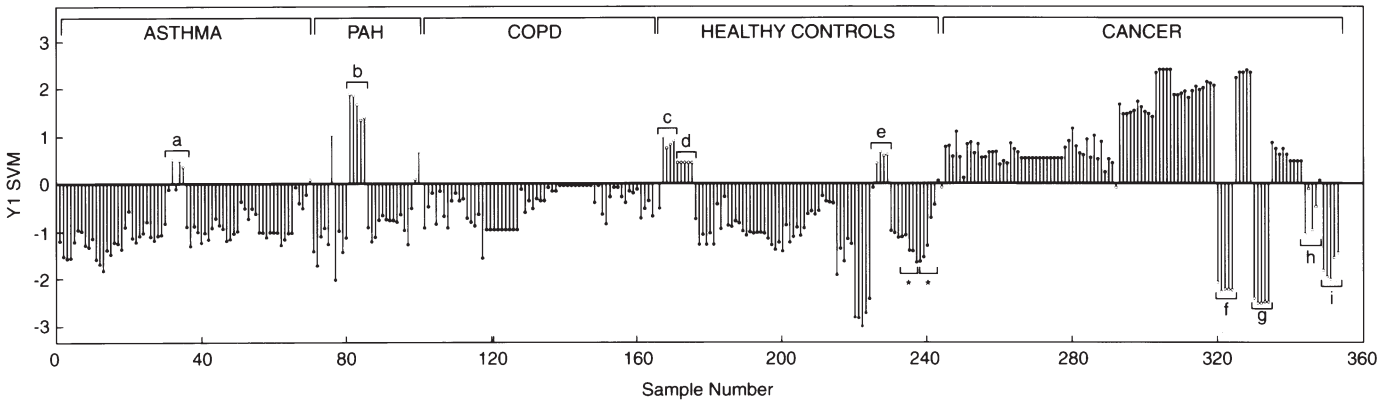

Figure 2. Support vector machine (SVM) classification for lung cancer. During classification, SVM calculates the distance of the unknown sample from the decision boundary in the model it has learned. In this graph, the margin for each breath sample is shown, with a positive value indicating classification of lung cancer (i.e., how far within the lung cancer boundary the sample falls). A minimum of five analyses performed on each individual's exhaled breath is shown. A negative value indicates a noncancer classification, with the value indicating how far outside of the lung cancer boundary the sample falls. The incorrect classification of a sample is identified by the *open circles* at the end of the line, and correct classification by *closed circles*. The majority of predictions were concordant (i.e., all five classifications of an individual the same in 92% of cases). Discordance occurred in 8% of cases. Assignment as cancer was predicted based on the predominant response (three or more of five) for that particular patient. Incorrect classification of lung cancer as noncancer is noted for two individuals with small cell carcinoma (f, g: all five analyses for each predict noncancer) and two individuals with relatively small primary lesions (h: four of five analyses predict noncancer; i: all five analyses for each predict noncancer). Incorrect classification of control subjects as cancer is noted for an individual with asthma with severe airflow limitation (a: three of five analyses cancer prediction), an individual with primary pulmonary hypertension (PAH; b: all five analyses predict cancer), and three healthy nonsmoking control subjects with no known lung disease (c, e: four analyses predict cancer; d: all five analyses predict cancer). *Indicates breath samples from two different individuals with lung cancer after curative resection of cancer.

TABLE 3. ACCURACY INDICES OF THE ELECTRONIC NOSE FOR DETECTION OF LUNG CANCER

Subgroup	Lung Cancer Present (*n*)	Lung Cancer Absent (*n*)	Sensitivity (*95% CI*)	Specificity (*95% CI*)	Positive Predictive Value (*95% CI*)	Negative Predictive Value (*95% CI*)
Positive exhaled breath test	10	5				
Negative exhaled breath test	4	57				
Total	14	62	71.4% (41.9–91.6) n = 10/14	91.9% (82.1–97.3) n = 57/62	66.6% (38.3–88.1) n = 10/15	93.4% (84–98.1) n = 57/61

Definition of abbreviation: CI = confidence interval.

electronic nose. Furthermore, a cancer prediction model based on the signature smellprints of exhaled breath of patients with lung cancer demonstrates an accuracy that suggests that the electronic nose could be a clinically useful tool in noninvasive lung cancer detection. As a part of routine medical health maintenance, screening for early detection of cancers has favorably influenced the survival of patients with cancer, including breast, colon, and prostate cancer (29). Unfortunately, widely applicable and effective screening for lung cancer, the leading cause of cancer death in the United States today, is not available (30).

Our results are comparable to previous studies evaluating lung cancer detection using gas chromatography and mass spectroscopy of exhaled gases (12, 13) or an electronic nose (15). The exhaled breath of humans contains a multitude of VOCs, many of which are in ambient air as well as endogenously produced, the most abundant being acetone, methanol, ethanol, propanol, and isoprene (5). Of these, pentane, isoprene, acetone, and benzene have been shown to be present in altered patterns in lung cancer exhaled gas samples analyzed by gas chromatography and mass spectroscopy (12–14). The results of gas chromatography and mass spectroscopy in our samples corroborate these findings and suggest that the electronic nose is capable of recognizing this distinct exhaled gas pattern. In comparison to gas chromatography and mass spectroscopy, the use of an electronic nose for detection of lung cancer offers several potential advantages, including ease of administration of the test and portability. In addition, this study brings a new method of sensor response analysis to the field of VOC-based exhaled breath detection of lung cancer.

The mechanisms leading to altered production of multiple VOCs in the exhaled breath of patients with lung cancer are likely related to increased oxidative events associated with the neoplastic process but are not related to cigarette smoking (31–34). The identification of patients after resection of lung cancer as noncancer provides further support that the source of the distinct pattern detectable by the electronic nose is the cancer. The high sensitivity for detection of non–small cell cancer suggests that our training set, which contained a preponderance of non–small cell cancer, may have selected for components distinct to this type of lung cancer. Finally, the distinct patterns observed in patients with lung cancer as opposed to those with chronic obstructive lung disease caused by α_1-AT deficiency, interstitial lung disease and patients evaluated in the validation set with COPD, asthma, and pulmonary arterial hypertension suggest that the findings could not be explained by the presence of airway inflammation or lung dysfunction alone.

Although not directly comparable to the context of this study, it is useful to compare the performance of the electronic nose to that of other noninvasive techniques used in detecting lung cancer. Contrast-enhanced spiral computed tomography has a sensitivity of 95 to 100% and a specificity reported between 58 and 93% (35). On the basis of a recent meta-analysis, positron emission tomography has 98.6% sensitivity and 77.8% specificity for identifying a malignant process in patients with pulmonary nodules or masses (36). In contrast to these tests, the analyses of exhaled breath for smellprints characteristic of bronchogenic cancer provide a potentially simple, noninvasive, and inexpensive screening tool.

On the other hand, a recent study using a quartz microbalance–based sensor evaluated 35 patients with active lung cancer, nine with resected disease and 18 healthy control subjects (15). Using partial least squares discriminant analysis as a supervised technique, the electronic nose had an overall accuracy of 90.3%. Furthermore, all of the patients with active lung cancer were correctly identified, five postsurgery patients were classified as control subjects and one healthy individual was classified as postsurgery. Our study results not only validate these previous observations but also add data from patients with chronic non-

TABLE 4. VOLATILE ORGANIC COMPOUNDS PRESENT IN THE EXHALED BREATH OF PATIENTS WITH LUNG CANCER

	Patient No.							
VOC (*ppbv*)	1	2	3	4	5	6	7	8
Isobutane	11	n.d.	5.6	13	n.d.	12	3.3	5.9
Methanol	63	n.d.	81	110	n.d	n.d.	n.d.	82
Ethanol	350	220	270	350	2160	1100	310	64
Acetone	150	190	870	240	370	260	220	270
Pentane	1	2	3	1	2	2	2	2
Isoprene	140	99	100	190	120	160	120	3
Isopropanol	270	1000	680	230	370	290	390	280
Dimethylsulfide	1.8	0.4	n.d.	0.7	3	1.9	1.1	2.4
Carbon disulfide	1.4	n.d.	1.6	3	n.d.	1.6	n.d.	n.d.
Benzene	3.45	n.d.	6.6	1.4	0.9	3.5	1.1	1.3
Toluene	6.4	4.6	3.2	1.9	4	6.4	4.5	3.2

Definition of abbreviations: n.d = not detected; ppbv = parts per billion volume.

neoplastic lung diseases. Collectively, these data suggest that biosensor analysis of the exhaled breath may be useful in the detection of lung cancer.

Still, several limitations of our study warrant comment. First, because many of the subjects in the current study had relatively advanced disease, further work needs to be done in a larger, more diverse population, which includes patients with less advanced disease and those with suspected lung neoplasms. Second, in the context that lung cancer was present in 18% of the validation patient cohort, the predictive values must be interpreted cautiously. Specifically, further study is needed to understand optimal strategies for using the electronic nose in population-based screening (where the negative predictive value would be expected to be higher because of much lower lung cancer prevalence) and in evaluating populations where the lung cancer frequency differs from that in our validation cohort.

Overall, in the context of previous studies documenting alterations in exhaled gases of patients with lung cancer (9–15), this study demonstrates the feasibility of clinical monitoring of VOCs in exhaled breath using a multisensor electronic nose as a relatively convenient and noninvasive test in patients with suspected lung cancer.

Conflict of Interest Statement: R.F.M. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; D.L. is the owner of Physiologic Measurement Systems LLC, which builds gas-sampling systems; O.D. is employed by Smiths Detection, which sponsored the study through the loan of the electronic nose detection device; T.B. is employed by Smiths Detection, which sponsored the study through loan of the electronic nose detection device; S.Z. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; P.J.M. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; T.M. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; C.J. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; J.K.S. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; J.P. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; J.D. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; R.A.D. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; S.C.E. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

References

1. Munoz BC, Steinthal G, Sunshine S. Conductive polymer-carbon black composites-based sensor arrays for use in an electronic nose. *Sensor Rev* 1999;19:300–305.
2. Kermany BG, Schiffman SS, Nagle HT. A novel method for reducing the dimensionality in a sensor array. *IEEE Trans Instr Meas* 1998;47: 728–741.
3. Kermany BG, Schiffman SS, Nagle HT. Using neural networks and genetic algorithms to enhance performance in an electronic nose. *IEEE Trans Biomed Eng* 1999;46:429–439.
4. Gardner JW, Bartlett PN. Electronic noses: principles and applications. Oxford, UK/New York: Oxford University Press; 1999.
5. Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. *Proc Natl Acad Sci USA* 1971;68:2374–2376.
6. Parry AD, Chadwick PR, Simon D, Oppenheim B, McCollum CN. Leg ulcer odour detection identifies beta-haemolytic streptococcal infection. *J Wound Care* 1995;4:404–406.
7. Lai SY, Deffenderfer OF, Hanson W, Phillips MP, Thaler ER. Identification of upper respiratory bacterial pathogens with the electronic nose. *Laryngoscope* 2002;112:975–979.
8. Hanson CW, Steinberger HA. The use of a novel electronic nose to determine the etiology of intrapulmonary infection [abstract]. *Anesthesiology* 1997;87:A269.
9. Pavlou AK, Magan N, McNulty C, Jones J, Sharp D, Brown J, Turner AP. Use of an electronic nose system for diagnoses of urinary tract infections. *Biosens Bioelectron* 2002;17:893–899.
10. Pavlou AK, Magan N, Sharp D, Brown J, Barr H, Turner AP. An intelligent rapid odour recognition model in discrimination of Helicobacter pylori and other gastroesophageal isolates in vitro. *Biosens Bioelectron* 2000;15:333–342.
11. Dalton P, Gelperin A, Preti G. Volatile metabolic monitoring of glycemic status in diabetes using electronic olfaction. *Diabetes Technol Ther* 2004;6:534–544.
12. Phillips M, Gleeson K, Hughes JM, Greenberg J, Cataneo RN, Baker L, McVay WP. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet* 1999;353:1930–1933.
13. Phillips M, Cataneo RN, Cummin AR, Gagliardi AJ, Gleeson K, Greenberg J, Maxfield RA, Rom WN. Detection of lung cancer with volatile markers in the breath. *Chest* 2003;123:2115–2123.
14. Gordon SM, Szidon JP, Krotoszynski BK, Gibbons RD, O'Neill HJ. Volatile organic compounds in exhaled air from patients with lung cancer. *Clin Chem* 1985;31:1278–1282.
15. Di Natale C, Macagnano A, Martinelli E, Paolesse R, D'Arcangelo G, Roscioni C, Finazzi-Agro A, D'Amico A. Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. *Biosens Bioelectron* 2003;18:1209–1218.
16. Machado R, Mazzone P, Laskowski D, Zheng S, Deffenderfer O, Burch T, Mekkhail T, Stoller J, Dweik R, Erzurum S. An electronic nose can diferentiate the exhaled breath of individuals with lung cancer [abstract]. *Am J Respir Crit Care Med* 2003;167:A190.
17. Murphy S, Sheffer A, Pauwels R. National Asthma Education and Prevention Program: highlights of the expert panel report II: guidelines for the diagnosis and management of asthma. Bethesda, MD: National Heart, Lung, and Blood Institute; 1997. NIH Publication No. 97-4051.
18. Rich S. Primary pulmonary hypertension: executive summary from the World Symposium on Pulmonary Hypertension. In: Rich S, editor. Geneva, Switzerland: World Health Organization; 1998.
19. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) workshop summary. *Am J Respir Crit Care Med* 2001;163:1256–1276.
20. Machado RF, Stoller JK, Laskowski D, Zheng S, Lupica JA, Dweik RA, Erzurum SC. Low levels of nitric oxide and carbon monoxide in alpha 1-antitrypsin deficiency. *J Appl Physiol* 2002;93:2038–2043.
21. Savitsky A, Golay MJE. Smoothing and differentiation of data by simplified least square procedures. *Anal Chem* 1964;36:1627–1639.
22. Wold S, Ebensen K, Geladi P. Principal component analysis. *Chemom Intell Lab Syst* 1987;2:37–52.
23. De Maesschalck R, Jouan-Rimbaud D, Massart DL. The Mahalanobis distance. *Chemom Intell Lab Syst* 2000;50:1–18.
24. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 2000;16:906–914.
25. Peng S, Xu Q, Ling XB, Peng X, Du W, Chen L. Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. *FEBS Lett* 2003;555:358–362.
26. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, Haussler D. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* 2000;97:262–267.
27. Kohlmann A, Schoch C, Schnittger S, Dugas M, Hiddemann W, Kern W, Haferlach T. Pediatric acute lymphoblastic leukemia (ALL) gene expression signatures classify an independent cohort of adult ALL patients. *Leukemia* 2004;18:63–71.
28. Belusov AI, Verkazov SA, von Frese J. Applicational aspects of support vector machines. *J Chemometrics* 2002;16(8–12):482–489.
29. Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer, 2003. *CA Cancer J Clin* 2003;53:27–43.
30. Bach PB, Niewoehner DE, Black WC. Screening for lung cancer: the guidelines. *Chest* 2003;123:83S–88S.
31. Feig DI, Reid TM, Loeb LA. Reactive oxygen species in tumorigenesis. *Cancer Res* 1994;54:1890s–1894s.
32. Dreher D, Junod AF. Role of oxygen free radicals in cancer development. *Eur J Cancer* 1996;32A:30–38.
33. Chung-man Ho J, Zheng S, Comhair SA, Farver C, Erzurum SC. Differential expression of manganese superoxide dismutase and catalase in lung cancer. *Cancer Res* 2001;61(23):8578–8585.
34. Kneepkens CM, Ferreira C, Lepage G, Roy CC. The hydrocarbon breath test in the study of lipid peroxidation: principles and practice. *Clin Invest Med* 1992;15:163–186.
35. Zhang M, Kono M. Solitary pulmonary nodules: evaluation of blood flow patterns with dynamic CT. *Radiology* 1997;205:471–478.
36. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA* 2001;285:914–924.