

Prediction of lung cancer using volatile biomarkers in breath¹

Michael Phillips^{a,b,*}, Nasser Altorki^c, John H.M. Austin^d, Robert B. Cameron^f, Renee N. Cataneo^a, Joel Greenberg^{a,**}, Robert Kloss^g, Roger A. Maxfield^e, Muhammad I. Munawar^a, Harvey I. Pass^h, Asif Rashidⁱ, William N. Rom^j and Peter Schmitt^k

^a*Menssana Research Inc, Fort Lee, NJ 07024-6510, USA*

^b*Department of Medicine, New York Medical College, Valhalla, NY., USA*

^c*Division of Thoracic Surgery, Department Of Cardiothoracic Surgery, Weill Medical College of Cornell University, New York, NY 10021, USA*

^d*Department of Radiology, Columbia University Medical Center, New York, NY 10032, USA*

^e*Department of Medicine, Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University Medical Center, New York, NY 10032, USA*

^f*Divisions of Thoracic Surgery and Surgical Oncology, UCLA School of Medicine and West Los Angeles VA Medical Center, Los Angeles, CA 90095-7313, USA*

^g*Danbury Hospital, Danbury, CT 06810, USA*

^h*Karmanos Cancer Institute, Wayne State University, 3990 John R, Suite 2102, Detroit MI 48201, USA*

ⁱ*Saint Vincent Catholic Medical Centers, Staten Island campus, Staten Island, NY 10310, USA*

^j*Division of Pulmonary and Critical Care Medicine, New York University Medical Center, New York, NY 10016, USA*

^k*Schmitt & Associates, 211 Warren St, Newark, NJ 07103, USA*

Abstract. *Background:* Normal metabolism generates several volatile organic compounds (VOCs) that are excreted in the breath (e.g. alkanes). In patients with lung cancer, induction of high-risk cytochrome p450 genotypes may accelerate catabolism of these VOCs, so that their altered abundance in breath may provide biomarkers of lung cancer.

Methods: VOCs in 1.0 L alveolar breath were analyzed in 193 subjects with primary lung cancer and 211 controls with a negative chest CT. Subjects were randomly assigned to a training set or to a prediction set in a 2:1 split. A fuzzy logic model of breath biomarkers of lung cancer was constructed in the training set and then tested in subjects in the prediction set by generating their typicality scores for lung cancer.

Results: Mean typicality scores employing a 16 VOC model were significantly higher in lung cancer patients than in the control group ($p < 0.0001$ in all TNM stages). The model predicted primary lung cancer with 84.6% sensitivity, 80.0% specificity, and 0.88 area under curve (AUC) of the receiver operating characteristic (ROC) curve. Predictive accuracy was similar in TNM stages 1 through 4, and was not affected by current or former tobacco smoking. The predictive model achieved near-maximal performance with six breath VOCs, and was progressively degraded by random classifiers. Predictions with fuzzy logic were consistently superior to multilinear analysis. If applied to a population with 2% prevalence of lung cancer, a screening breath test would have a negative predictive value of 0.985 and a positive predictive value of 0.163 (true positive rate = 0.277, false positive rate = 0.029).

¹Prior disclosure: A report of this study was presented at the scientific meeting of the American Society of Clinical Oncology, Orlando, FL, May 16, 2005.

*Corresponding author: Michael Phillips MD, FACP, Menssana

Research, Inc. 1 Horizon Road, Suite 1415, Fort Lee, NJ 07024, USA.
Tel./Fax: +1 201 886 7004; E-mail: mphilips@menssana-research.com.

**Deceased.

Conclusions: A two-minute breath test predicted lung cancer with accuracy comparable to screening CT of chest. The accuracy of the test was not affected by TNM stage of disease or tobacco smoking. Alterations in breath VOCs in lung cancer were consistent with a non-linear pathophysiologic process, such as an off-on switch controlling high-risk cytochrome p450 activity. Further research is needed to determine if detection of lung cancer with this test will reduce mortality.

1. Introduction

Primary carcinoma of the lung is the major cause of cancer death in the United States.

Patients who are diagnosed when the tumor is still confined to a primary site have the best prospects for survival: five years after diagnosis, fewer than 5% patients with stage 4 lung cancer are still alive, compared to 60–80% patients with stage 1 disease [1]. Researchers have therefore sought new ways to detect early-stage lung cancer, such as microanalysis of volatile organic compounds (VOCs) in the breath. This technology, pioneered by Linus Pauling in 1971 [2], was reported to identify candidate biomarkers of lung cancer during the 1980s [3,4]. Members of our group have previously reported abnormal patterns of breath VOCs in patients with early-stage primary lung cancer [5,6], as well as in other diseases, including heart transplant rejection [7] (approved for clinical use by the Food & Drug Administration), breast cancer [8], unstable angina [9], preeclampsia of pregnancy [10], and diabetes mellitus [11].

Our initial study of primary lung cancer empirically identified a set of 22 VOCs in breath as biomarkers of disease. A mathematical model employing these VOCs, principally alkanes, alkane derivatives and benzene derivatives, was sensitive and specific for primary lung cancer and equally accurate in early as well as in advanced disease [5]. Our second report incorporated incremental improvements in technique: we employed a rational set of breath markers of oxidative stress instead of the empirically identified VOCs, and a more sensitive analytical method that reduced breath collection time from 5.0 min to 2.0 min. That study confirmed the accuracy of breath testing for primary lung cancer: a model employing nine alkanes and methylated alkanes was sensitive, specific, and equally accurate in all TNM stages of disease [6]. The abundance of the breath biomarkers was significantly lower in patients with primary lung cancer than in normal controls, suggesting a biological mechanism: VOC products of oxidative stress may have undergone accelerated catabolism by cytochrome P450 mixed oxidases whose activity was induced during the pathogenesis of lung cancer [12–14].

This report incorporates additional improvements in analytical methodology and experimental design. We

enrolled a sufficient number of subjects to develop a model in one group and test its predictive power in a second group. Cancer-free controls were validated with negative chest CT imaging, and the selectivity of the analytical method was increased in order to identify a greater number of VOCs in breath. Also, we analyzed breath VOC data with fuzzy logic, an advanced technique of multivariate analysis. However, a disadvantage of improved analytical methodology was that breath VOC data obtained in previous studies could not be employed over again.

Multivariate analysis of several markers in combination can frequently predict disease with greater accuracy than a single marker employed alone. Fuzzy logic analysis of multiple markers is a powerful statistical method that has been successfully employed to predict a wide variety of disorders [15], including breast cancer [16], prostate cancer [17], cardiac arrhythmias [18], Tourette's syndrome [19] and suicide risk [20]. Fuzzy logic differs from conventional statistical methods of multivariate analysis because it employs ranges of values and their relationships to one another, instead of exact values. For example, when employed in breath testing, fuzzy logic might identify an abnormally high abundance of one breath VOC as a marker of disease, but only if it is observed in combination with an abnormally low value of another breath VOC. We report here a multi-center study that employed fuzzy logic analysis of volatile biomarkers in the breath in order to predict the presence or absence of primary lung cancer.

2. Methods

2.1. Study design

An overview is shown in Fig. 1. Breath VOCs were analyzed in patients with untreated primary lung cancer and in controls with no evidence of cancer on spiral computed tomography (CT) of the chest. Subjects were randomly assigned to a training set in order to construct a fuzzy logic model of breath VOC markers of lung cancer, or to a prediction set where the model was tested as a predictor of disease.

Table 1
Patient characteristics. This table shows the demographics of the primary lung cancer patients and the cancer-free controls. Sex distribution, age and pack-years of tobacco smoking were not significantly different in the two groups (*chi-squared test, **2-tailed t-test)

	Primary lung cancer	Controls	
No.	193	211	
Non-small cell cancer	189		
Small cell cancer	4		
Sex (M/F)	96/97	106/105	N.S.*
Mean age in yrs (SD)	66.0 (11.4)	67.5 (5.5)	N.S.**
Tobacco smoking history			
Pack-years of cigarette smoking (SD)	41.9 (36.2)	43.9 (29.4)	N.S.**
Smoking activity:			
Current smokers	33	78	
Non-smokers	24	0	
Former-smokers	134	133	
Activity not recorded	2	0	

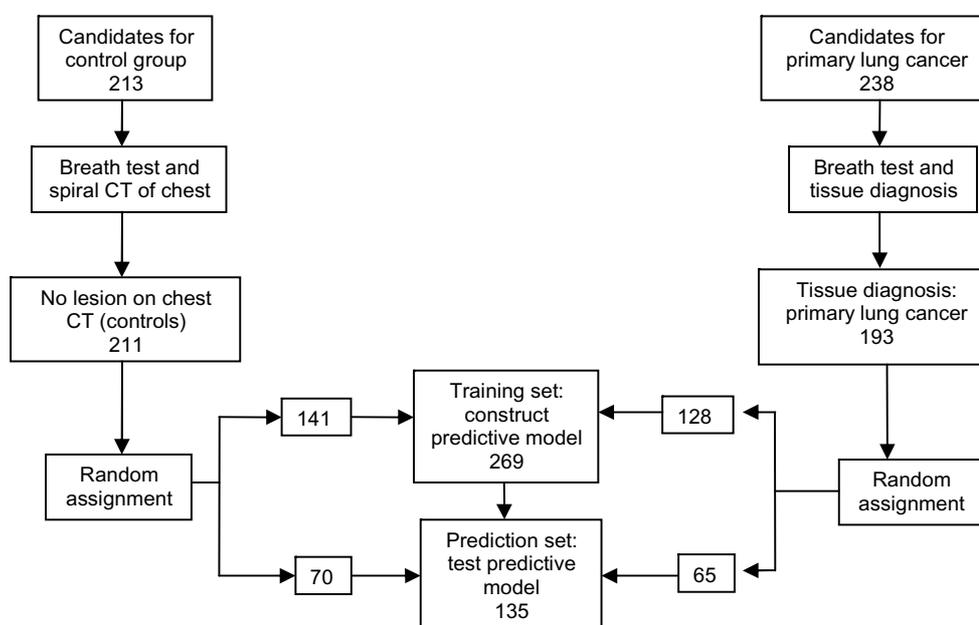


Fig. 1. Design of the study. Candidates for the control group were recruited from a group of asymptomatic smokers aged ≥ 60 yr, who underwent spiral CT of the chest in a study of primary screening for lung cancer. Only those with no detectable lesions on chest CT were included in the control group. Candidates for the primary lung cancer group were recruited from patients who had been referred for a tissue diagnosis because of a suspicious lesion on chest imaging. All donated a breath sample for analysis of volatile organic compounds (VOCs). Subjects were randomly selected from the control group and the primary lung cancer group in order to achieve an approximately 2:1 split between the training set and the prediction set. The fuzzy logic software constructed a predictive model for primary lung cancer employing the breath VOCs of subjects in the training set, and this predictive model was tested for its ability to predict the presence or absence of lung cancer in subjects in the prediction set.

2.2. Patient recruitment

The research was approved by the Institutional Review Boards of all collaborating institutions; all subjects gave their signed informed consent to participate.

Candidates for the primary lung cancer group were recruited from patients referred to a pulmonary or sur-

gical service for tissue diagnosis following the first occurrence of a suspicious lesion observed on thoracic imaging. Patients with a history of treated lung cancer were excluded. Surgical biopsy was performed in 238 candidates recruited at Harper Hospital (117), Weill Medical College of Cornell University (56), University of California Los Angeles Medical Center (30), New

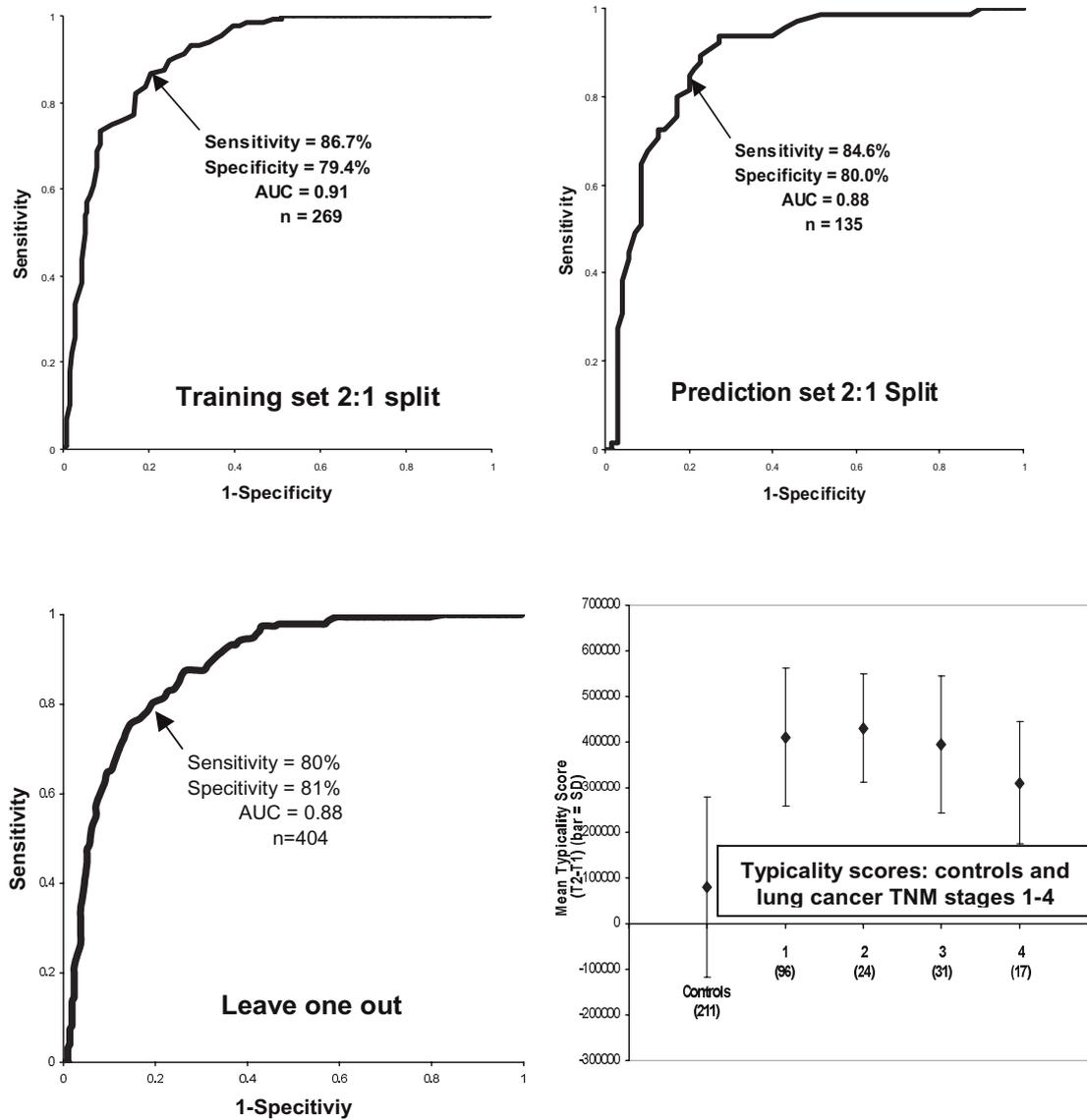


Fig. 2. *Breath test results – training and prediction sets.* In the training set, fuzzy functions were derived from the breath VOCs to construct two typicality matrices, one for controls and the other for primary lung cancer patients. These matrices generated two numerical values for each subject: T_1 , the typicality for no disease, and T_2 , the typicality for primary lung cancer. The value of $T_2 - T_1$ was employed as a predictor of primary lung cancer, and its accuracy is shown in the contour of the receiver operating characteristic (ROC) curve and its area under curve (AUC). The arrow in each ROC curve indicates where the sum of sensitivity plus specificity was maximal. If a test delivers results no better than chance alone, then the ROC curve is a 45° straight line commencing from the origin, and $AUC = 0.5$. Conversely, if a test delivers completely accurate results with no false positives or false negatives, then the ROC curve is a right angle with its apex at the top left of the panel, and $AUC = 1.0$. A test with $AUC > 0.9$ is usually deemed to exhibit excellent performance. *Top left panel. Training set 2:1 split:* This figure shows the ROC curve obtained with fuzzy logic analysis of breath VOCs in the training set comprising two thirds of the primary lung cancer patients and controls. *Top right panel. Prediction set 2:1 split:* This figure shows the ROC curve obtained by employing each subject's typicality value of $T_2 - T_1$ as a predictor of primary lung cancer in the prediction set comprising one third of subjects. *Bottom left panel. Prediction set leave-one-out:* This figure shows the ROC curve obtained with a training set comprising $n-1$ subjects and the prediction set comprising one subject. The process is iterated n times to predict the outcome in all subjects. The ROC prediction set curve appears smoother than in the top right panel because it includes points for all subjects instead of one third. VOCs were restricted to the 16 shown in Table 2. *Bottom right panel. Prediction set leave-one-out:* This figure shows the mean typicality score of $T_2 - T_1$ in the prediction set stratified according to TNM stage of disease. There were no significant differences between the typicality scores of the patients with different TNM stages of lung cancer; the mean typicality score in the control group was significantly lower than in the lung cancer patients ($p < 0.0001$ in all TNM stages, analysis of variance and post hoc Newman-Keuls test).

Table 2

Major VOC identifiers of primary lung cancer in breath. This table displays the breath VOCs identified as biomarkers of lung cancer by multilinear regression and fuzzy logic. VOCs are ranked according to their cumulative contribution to the AUC of the ROC curves in the training and prediction sets as shown in Fig. 4, and are numbered in the same sequence. Multilinear regression required 11 VOCs in order to achieve a model with maximal performance, and fuzzy logic approached a maximal performance plateau with 16 VOCs

no.	Chemical structure	AUC of ROC curve			
		Multilinear		Fuzzy logic	
		Train	Predict	Train	Predict
1	1,5,9-Cyclododecatriene, 1,5,9-trimethyl-	0.57	0.57	0.06	0.07
2	Pentan-1,3-dioldiisobutyrate, 2,2,4-trimethyl	0.69	0.70	0.71	0.73
3	Benzoic acid, 4-ethoxy-, ethyl ester	0.69	0.69	0.76	0.79
4	Propanoic acid, 2-methyl-, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester	0.77	0.76	0.81	0.84
5	10,11-dihydro-5H-dibenz-(B,F)-azepine	0.77	0.76	0.83	0.87
6	2,5-Cyclohexadiene-1,4-dione, 2,6-bis(1,1-dimethylethyl)-	0.79	0.80	0.83	0.87
7	Benzene, 1,1-oxybis-	0.79	0.78	0.85	0.86
8	Furan, 2,5-dimethyl-	0.80	0.77	0.87	0.86
9	1,1-Biphenyl, 2,2-diethyl-	0.81	0.76	0.87	0.87
10	3-Pentanone, 2,4-dimethyl-	0.81	0.78	0.87	0.87
11	trans-Caryophyllene	0.81	0.78	0.87	0.87
12	1H-Indene, 2,3-dihydro-1,1,3-trimethyl-3-phenyl-			0.87	0.86
13	1-Propanol			0.86	0.86
14	Decane, 4-methyl-			0.88	0.86
15	1,2-Benzenedicarboxylic acid, diethyl ester			0.88	0.86
16	2,4-Hexadiene, 2,5-dimethyl-			0.91	0.88

York University Medical Center (20), Columbia University Medical Center (11), and Danbury Hospital (4). All patients having resection of primary lung cancer had systematic mediastinal lymph node dissection or lymph node sampling. Admission to the primary lung cancer group was based on the reported histopathology of a patient's biopsy specimens. The pathologic stage of the disease was determined according to the International TNM (tumor, node, metastasis) staging system for lung cancer by examination of the pathological tumor specimen and resected lymph nodes.

Candidates for the control group were recruited from 213 asymptomatic smokers aged 60 and over who participated in the Early Lung Cancer Action Program (ELCAP) for primary detection of lung cancer [21], and underwent spiral CT imaging of the chest at Columbia University Medical Center. Subjects were entered into the control group if imaging revealed no pulmonary lesion suggestive of cancer.

Post-operative patients: A subgroup of the primary lung cancer group ($n = 80$) had a second breath test following surgical resection of the tumor, and the predictions of the fuzzy logic model were tested.

2.3. Masking procedures

At all sites, tissue samples were interpreted by pathologists without knowledge of the breath test results. Analyses of breath VOCs were performed by RNC without knowledge of the pathological findings.

2.4. Breath collection and assay

The method has been previously described [22]. In summary, subjects breathed normally through the disposable mouthpiece of a portable breath collection apparatus for 2.0 min and the VOCs in 1.0 L alveolar breath and 1.0 L room air were captured onto separate sorbent traps. Light flap valves in the mouthpiece presented low resistance to respiration so that samples were collected without discomfort to patients, including the elderly. VOCs captured in the sorbent traps were analyzed in the laboratory by automated thermal desorption, gas chromatography and mass spectroscopy. The alveolar gradient of each VOC was determined as its abundance in breath minus its abundance in room air.

2.5. Statistical methods

Data analysis was performed by JG and PS. Subjects were randomly assigned to a training set or a prediction set in a ratio of approximately 2:1. Candidate breath VOC biomarkers of lung cancer (present in the breath of at least 200 subjects) were initially selected from the training set by comparing their abundance in the primary lung cancer and control groups with t-tests. VOCs exhibiting a significant difference between the two groups ($p < 0.05$) were then entered into multivariate analysis with fuzzy logic and with forward stepwise multi-linear regression.

Table 3

Patients excluded from training set and prediction set: 45/238 primary lung cancer candidates were excluded from entry into model building in the training set or model testing in the prediction set (Fig. 1) for the following indications

Primary lung cancer without demographic and/or other clinical data	29
Metastatic cancer	7
Mesothelioma	4
Benign tumor	3
Recurrent tumor	2

2.6. Fuzzy logic analysis of training set

Interrelation Miner (SystAim[®], Zürich, Switzerland) was employed. Fuzzy functions were constructed for the candidate breath VOCs in order to create one typicality matrix for controls and another typicality matrix for primary lung cancer patients.

2.7. Fuzzy logic analysis of prediction set

The typicality matrices were employed to generate two numerical values from the breath VOCs: T_1 , the typicality for no disease, and T_2 , the typicality for primary lung cancer. The value of $T_2 - T_1$ was employed as a predictor of primary lung cancer, and the accuracy of prediction was displayed in a receiver operating characteristic (ROC) curve. As an additional cross validation, the process was repeated in all subjects combined, employing a leave-one-out procedure. $n-1$ subjects were entered into a training set to construct the typicality matrices, and the outcome was predicted in the remaining one subject. The procedure was iterated n times, in order to predict the outcome in every subject.

2.8. Linear regression analysis

Data were also analyzed with forward stepwise multi-linear regression. A model was constructed with the same training set data, and evaluated in the same prediction set that was employed for fuzzy logic.

2.9. Random classifiers

Class labels (cancer and controls) were randomly reversed for VOCs in the training and prediction sets in order to determine the changes in predictive capabilities of fuzzy logic modeling of data.

3. Results

3.1. Human subjects

No subject reported any adverse effects of donating a breath sample. Characteristics of subjects in the primary lung cancer and control groups are shown in Table 1. There were no significant differences between the two groups in their mean age, sex distribution or pack-years of smoking history. Controls were excluded from analysis if CT imaging demonstrated a lesion consistent with cancer, and primary lung cancer candidates were excluded from analysis if another pathology (e.g. metastatic cancer) was demonstrated or if TNM staging information or other demographic data was not available (Table 3).

3.2. Candidate biomarkers of lung cancer

35 candidate breath VOCs were identified in the training set data. Fuzzy logic selected 16 of these VOCs for inclusion in the typicality matrices, and multi-linear regression selected 11 of these VOCs and no others (Table 2).

4. Prediction of lung cancer

4.1. Controls and primary lung cancer

The mean typicality scores were significantly higher in the lung cancer patients than in the control group ($p < 0.0001$ in all TNM stages, analysis of variance and post hoc Newman-Keuls test). Receiver operating characteristic (ROC) curves displaying the results of the breath test in the training set and prediction sets are shown in Fig. 2. The breath test predicted primary lung cancer with 84.6% sensitivity and 80.0% specificity where the sum of sensitivity and specificity was maximal; the area under curve (AUC) of the ROC curve was 0.88. Figure 8 shows a histogram of the distribution of cancer patients and controls as a function of the typicality scores in the training set and prediction sets. Cross validation using all samples and employing a leave-one-out method was performed first with reselection of VOCs anew in every step. This model selected 22 VOCs and the predicted AUC of the ROC curve was 0.89, with 91.2% sensitivity and 73.5% specificity.

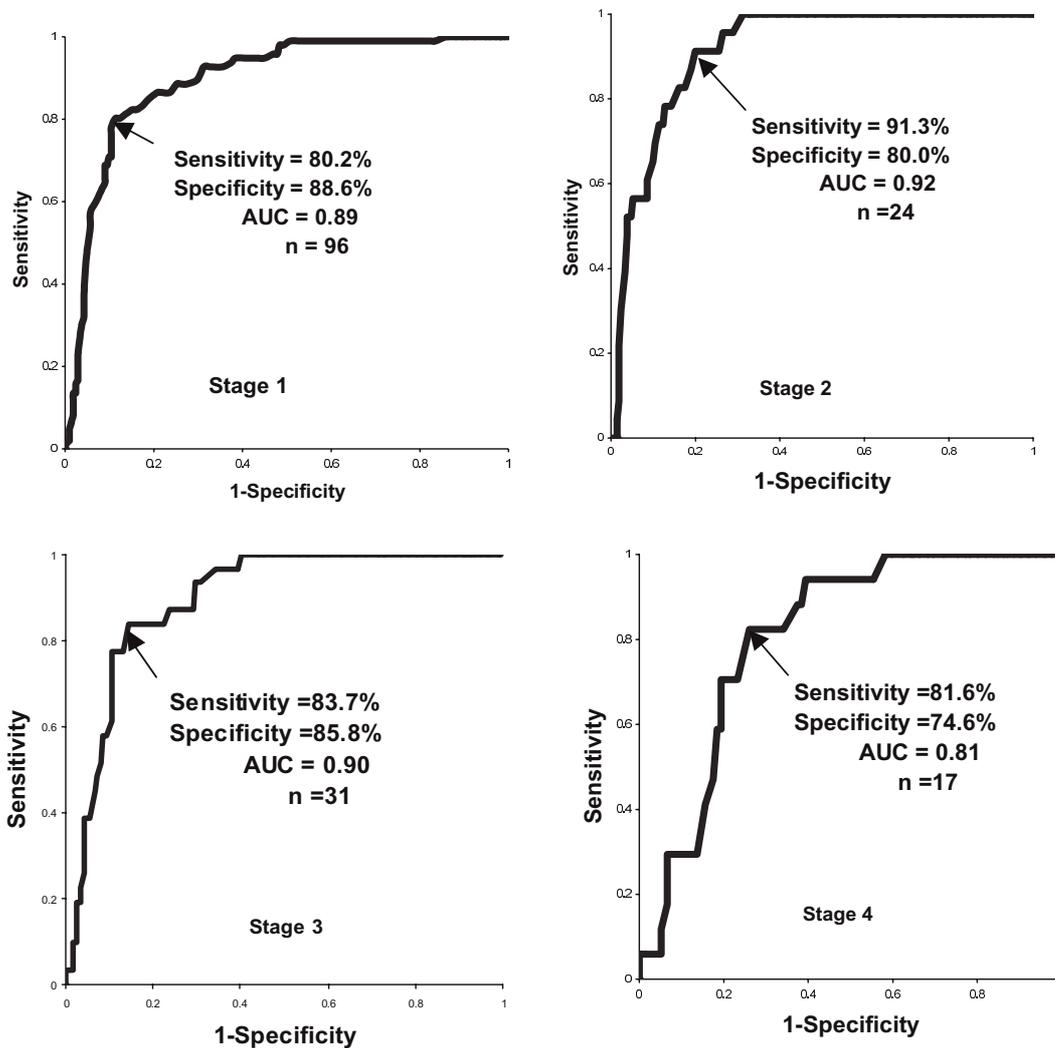


Fig. 3. Prediction of primary lung cancer stratified by TNM stage. This figure shows the predictions of the fuzzy logic model in all subjects employing the leave-one-out predictions (shown in Fig. 2) stratified according to TNM stage of lung cancer. Staging data was available for 168 patients. The arrow in each panel indicates the point on the ROC curve where the sum of sensitivity plus specificity was maximal. These ROC curves displayed similar accuracy of prediction in all TNM stages.

4.2. Different TNM stages of primary lung cancer

TNM staging data was available in 168/193 patients. The mean typicality scores stratified according to TNM stage are shown in Fig. 2 and the respective ROC curves are shown in Fig. 3. There were no significant differences between the predicted mean typicality scores in TNM stages 1 through 4.

4.3. Post-operative patients

The predictive model was positive for primary lung cancer in 77/80 (96.3%) subjects following surgical resection of the tumor.

4.4. Primary lung cancer candidates excluded from training and prediction sets

The predictive model was positive for primary lung cancer in 45/45 (100%) subjects, including 31/45 subjects with primary or recurrent lung cancer and 14/45 patients with other pathologies (metastatic cancer, mesothelioma, or benign tumors).

4.5. Effect of the number of VOCs on model performance

Figure 4 displays the AUC of the training and prediction ROC curves as a function of the number of

VOCs in the model. AUC values increased as more breath VOCs were entered into the model, nearing a maximum with six VOCs before leveling to a plateau. AUC values were consistently higher with fuzzy logic than with linear regression analysis. The maximal AUC of the prediction ROC curve was 0.88 with fuzzy logic (16 VOCs) and 0.78 with linear regression analysis (11 VOCs).

4.6. *Effect of random classifiers*

Figure 5 displays the effect of randomly reversing the assignment of VOC data between the cancer and the control groups and the resulting progressive degradation of the fuzzy logic model with random classifiers.

4.7. *Effect of tobacco smoking*

Table 1 demonstrates no significant difference between total exposure to tobacco products in controls and cancer patients as shown by pack-years of smoking history. Figure 6 displays the lack of effect of tobacco smoking on the accuracy of the predictive model. The AUC of the ROC curves in current smokers and in former smokers were similar to the results obtained in the prediction set for all subjects.

4.8. *Effect of multiple random data splits*

We employed the strategy of Michiels et al to evaluate the stability of the molecular signature of breath biomarkers of lung cancer by using 20 multiple random sets of unique training and prediction sets (Fig. 7) [23].

5. Discussion

The main finding of this study was that a combination of breath VOCs accurately predicted lung cancer. The accuracy of the breath test was comparable to screening chest CT. Sone et al reported a population screening study of 5483 subjects in which mobile low-dose spiral chest CT detected surgically confirmed lung cancer with 55% sensitivity and 95% specificity [24]. At the point on the prediction set ROC curve (Fig. 2) where the breath test detected lung cancer with 55% sensitivity, its specificity was 91%.

In addition, the breath test was equally sensitive and specific in early-stage lung cancer as in advanced metastatic disease, and the accuracy of the test was not affected by former or current tobacco smoking.

Despite differences in analytical methods and patient populations, the findings of this study were consistent with our previous reports of breath testing for lung cancer [5,6]. The major breath biomarkers were mainly alkane derivatives in all studies. Also, the abundance of most of these VOCs was decreased in patients with lung cancer, and the predictive accuracy of the breath test was similar in all TNM stages of disease.

The results illustrated differences between the pathophysiology of tumor markers in serum compared to volatile biomarkers of lung cancer in the breath.

5.1. *Molecular weight*

Tumor markers in serum are generally macromolecules with a molecular weight of several kilodaltons [25,26]; in contrast, breath biomarkers of lung cancer observed in this study were small molecules with molecular weights less than 600.

5.2. *Effect of tumor mass*

Some serum tumor markers increase in abundance with the size of the tumor [27]. In contrast, the predictive value of breath biomarkers of lung cancer (i.e., their mean typicality scores) did not vary significantly with the TNM stage of disease. Since typicality scores were similar in localized as well as in metastatic disease, an elevated typicality score might possibly precede the appearance of a macroscopically detectable lesion.

5.3. *Decreased vs. increased abundance*

Unlike serum tumor markers, the abundance of most breath biomarkers of lung cancer was decreased rather than increased.

5.4. *Effect of surgery*

The concentration of a serum tumor marker is generally reduced by excision of the cancer, e.g., prostate specific antigen is reduced by ablation of the prostate [28]. However, the predictions of the breath test were unchanged in the majority of post-operative patients following thoracotomy and resection of the pulmonary lesions.

Taken together, these findings are consistent with the hypothesis that volatile markers of primary lung cancer in breath emanate from other tissues, and not from the tumor itself. We have previously proposed

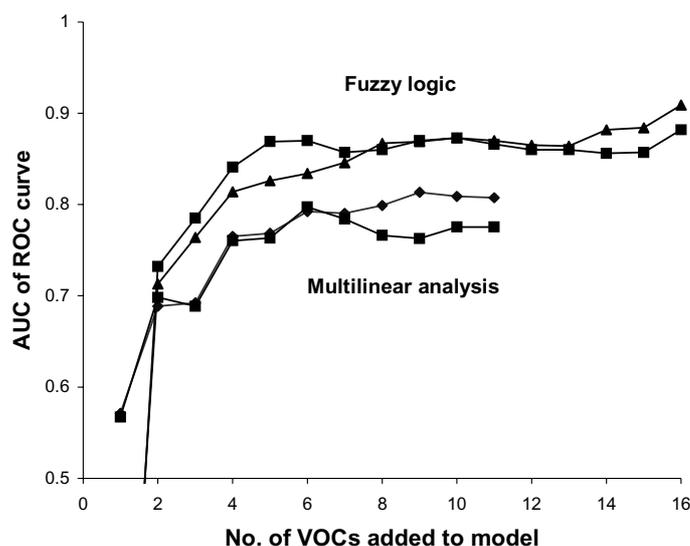


Fig. 4. Effect of the number of VOCs on model performance: This figure displays the AUC of the ROC curve as a function of the number of VOCs in the model (2:1 data split; triangles = fuzzy logic training set, diamonds = multilinear analysis training set; rectangles = prediction set). The maximal AUC values of the prediction set ROC curve were 0.78 employing forward stepwise multilinear analysis with 11 VOCs, and 0.88 employing fuzzy logic with 16 VOCs. The breath test achieved near maximal predictive accuracy with six VOCs in both methods. Addition of more VOCs to the model did not greatly improve the results, but the addition of redundancy to the model probably ensures that its performance will be more resistant to future systematic changes and random fluctuations. Results with fuzzy logic were consistently superior to multilinear analysis, consistent with an underlying non-linearity in the pathophysiologic processes underlying the model.

a pathophysiologic model that provides a rational basis for these observations [6]: Oxidative stress occurs in all tissues of the body and reactive oxygen species leak from the mitochondria, peroxidate polyunsaturated fatty acids in cell membranes, and generate volatile alkanes and methylated alkanes that are excreted in the breath [22,29]. These VOCs are catabolized by highly inducible cytochrome P450 mixed oxidases [12]. Polycyclic aromatic hydrocarbons in tobacco smoke induce CYP1A1, CYP1A2 and possibly CYP2E1. The risk of lung cancer is increased in individuals who carry high-risk genotypes for CYP2E1 and CYP1A1. Cytochrome P450 activity is predominantly hepatic, though CYP1A1 is primarily an extrahepatic enzyme found in lung and placenta [13,14]. We propose that induction of cytochrome P450 genotypes in high-risk individuals simultaneously increases the risk of lung cancer and accelerates catabolism of the volatile products of oxidative stress, thereby reducing their abundance in breath. This hypothesis is consistent with the unique features of breath tumor markers described above. Breath VOCs may possibly be altered before the appearance of macroscopically detectable lesions because induction of high risk cytochrome P450 phenotypes precedes the onset of neoplasia. Also, this hypothesis accounts for the lack of effect of surgical resection of lung lesions on breath biomarkers of lung

cancer, because the enzymes responsible for accelerated catabolism of alkanes are located in tissues remote from the tumor. The predictive model was also positive for primary lung cancer in 45/45 primary lung cancer candidates excluded from training and prediction sets (Table 3). This finding was consistent with the majority of the diagnoses (29/45 primary lung cancer without TNM staging and 2/45 with a recurrent tumor). However, 14/45 patients were excluded because of conditions other than primary lung cancer (metastatic cancer, mesothelioma, and benign tumors) (Table 3), demonstrating that the breath test may become positive in other conditions associated with abnormal chest imaging. This finding merits further evaluation in future studies.

Multivariate analysis of breath VOC data was required to identify the biomarkers of lung cancer because induction of polymorphous cytochrome P450 enzymes affects the abundance of several different compounds. We compared different techniques of multivariate analysis including discriminant analysis, logistic regression, pattern recognition analysis, neural networks and fuzzy logic. All generated accurate predictive models of lung cancer, but results with fuzzy logic were consistently superior to multilinear analysis (Fig. 4). This is consistent with a non-linear pathophysiologic process underlying the development of lung cancer e.g. the alteration in breath VOCs in lung cancer may have re-

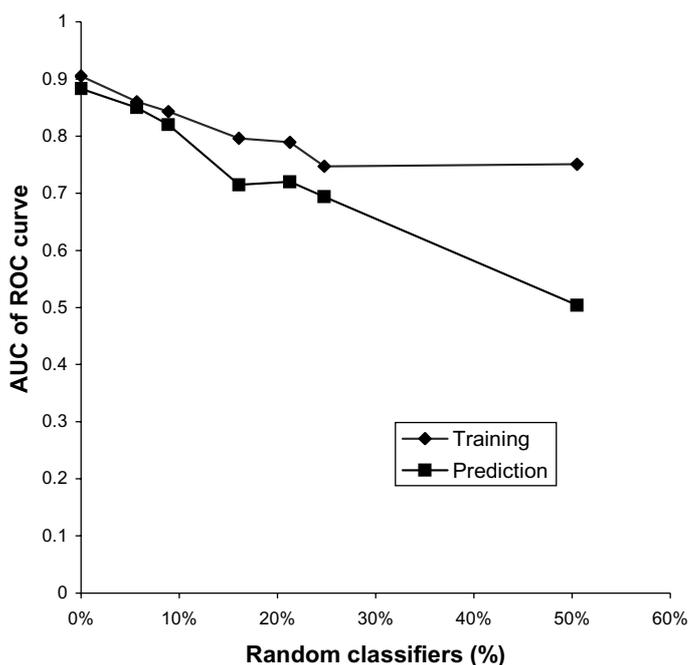


Fig. 5. Effect of random classifiers on fuzzy logic performance: This figure displays the effects of randomly reversing the assignment of VOC data between the cancer and the control groups. The AUC of the ROC curve degraded approximately 5% for every 10% of random classifier changes. When the classification was completely random (i.e. 50% of the classifiers were reversed), the AUC of the training set ROC was approximately 0.7, demonstrating that fuzzy logic overfitted the data by finding an apparent pattern within random data. However, this model had no predictive value because the AUC of the prediction set ROC fell to 0.5 (i.e. no better than random prediction). The addition of random classifiers progressively degraded the predictive model, which was consistent with the conclusion that the non-degraded 16 VOC model was an accurate predictor of lung cancer, and not the chance consequence of a fortuitous statistical model.

sulted from off-on switching of cytochrome P450 enzyme activity rather than a progressive continuum of increased activity.

Fuzzy logic is a powerful problem-solving methodology that recognizes more than simple true and false values by allowing propositions to be represented with degrees of truthfulness and falsehood. For example, the statement, today is sunny, might be 100% true if there are no clouds, 80% true if there are a few clouds, and 0% true if it rains all day. Computer software employing fuzzy logic is employed widely in industry and increasingly in clinical medicine because it resembles human decision making with its ability to derive precise solutions from approximate data. Others have also employed fuzzy logic to detect lung cancer by combining the contributions of different tumor markers in serum [30].

Typicality is a measure of how well a given set of measurements for one particular patient agrees with a list of predefined combinations of measurements in health or disease. Fuzzy logic was employed to construct a separate typicality matrix for controls and for patients with lung cancer. First, all of the alveolar gra-

dients observed in a specific VOC were assigned into separate classes e.g. 1 to 5 according to their value (low = 1, high = 5), and the number of subjects in each class was counted. This process was repeated for all VOCs that were entered into the model. The resulting matrix is termed a typicality matrix because it reflects the typical distribution of alveolar gradients of a number of VOCs in a particular group of subjects. Each subject's breath test results were then evaluated in order to determine two values, T_1 and T_2 , which varied with how well their distribution of alveolar gradients accorded with the typicality matrix for controls and for lung cancer respectively. The higher the value of T_1 or T_2 in a particular subject, the more typical is the breath test result for the control group or for lung cancer respectively. Both T_1 and T_2 are always positive. If $T_2 - T_1$ is positive, then the subject is more likely to belong to the lung cancer group; conversely, if $T_2 - T_1$ is negative then the subject is more likely to belong to the control group. The value of $T_2 - T_1$ may be regarded as an index of the likelihood of lung cancer, and employed in the construction of an ROC curve.

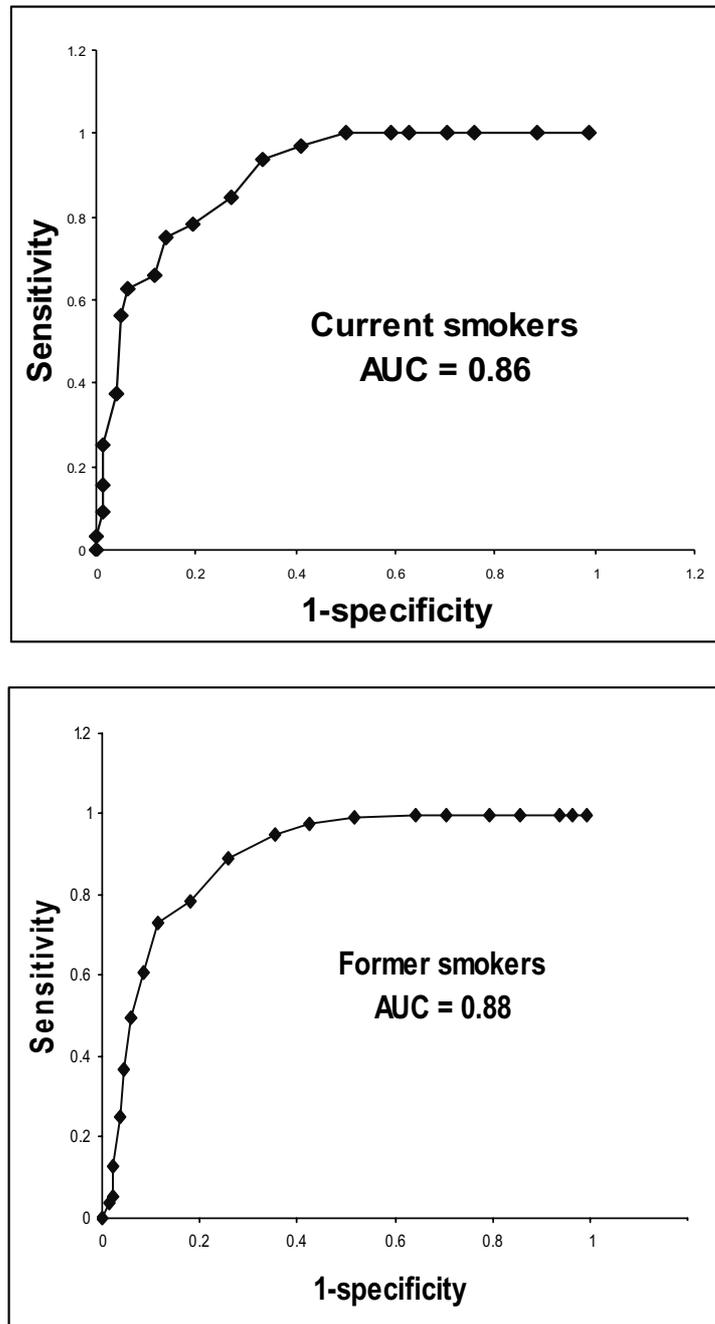


Fig. 6. Effect of tobacco smoking on accuracy of the predictive model: The predictive model was evaluated separately in current smokers (upper panel) and in former smokers (lower panel). The AUC of both ROC curves was similar to the results obtained in the prediction set for all subjects (Fig. 2), demonstrating that the accuracy of the predictive model was not skewed by current smoking or a history of smoking.

In addition, the finding that the predictive accuracy of the breath test progressively declined with the introduction of random classifiers (Fig. 5) supports the conclusion that the fuzzy logic model accurately predicted lung cancer, and was not the chance consequence

of a fortuitous statistical model. The similarity of the ROC curves with multiple random splits into training and prediction sets is consistent with a stable molecular signature of the breath biomarkers of lung cancer [23].

The expected performance of a screening test de-

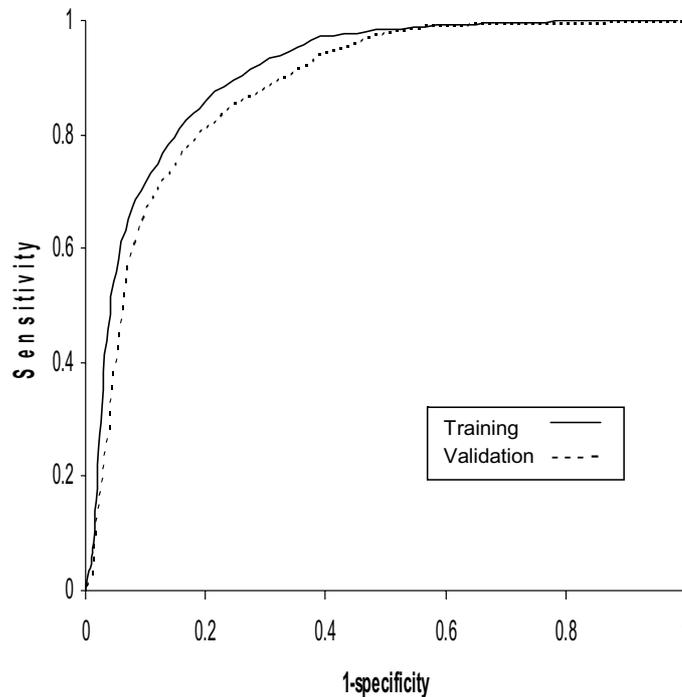


Fig. 7. Effect of multiple random data splits on accuracy of the predictive model: The data set was randomly split in a 2:1 ratio to generate 20 unique training sets and prediction sets. The mean AUCs of the training and prediction ROC curves were 0.907 (SD = 0.009) and 0.885 (SD = 0.025) respectively. The similarity of the resulting ROC curves was consistent with a stable molecular signature of the breath biomarkers of lung cancer, and exhibits the robustness of the overall approach.

depends not only on the performance of the test in diseased and non-diseased subjects, but also on the prevalence of primary lung cancer (1.7% to 2.7% in smokers aged over 50 yrs and 60 yrs respectively [31,32]). Performance is typically assessed by two statistics that describe the accuracy of the classification rule: the true positive rate (TPR) (sensitivity), and the false positive rate (FPR) (one minus specificity). Corresponding values of TPR and FPR may be selected from any point on the prediction set ROC curve shown in Fig. 2. However, FPR relates to the potential harms of a test and TPR relates to its potential benefits, so the selected combination of FPR and TPR should ensure that benefits will probably outweigh harms in a future study. For mammography, reasonable target values have been cited as FPR = 0.01 and TPR = 0.80 [33]. We therefore selected a point on the prediction set ROC curve where the test exhibited near-maximal specificity (FPR = 0.029 and TPR = 0.277). Assuming a prevalence of 2% in $N = 1000$ smokers aged over 50 yrs, our proposed screening breath test would have a negative predictive value (NPV) of 0.985 and a positive predictive value (PPV) of 0.163. In practice, individuals with a negative breath test result (the majority of the screened pop-

ulation) could be reassured that their risk of primary lung cancer is low. Since the NPV of the breath test is high, this group would require no further testing with diagnostic radiation. This could potentially reduce the number of screening chest CT studies performed, thereby reducing the number of unnecessary invasive studies that follow false-positive findings on chest CT screening as currently practiced. Conversely, the low FPR of the test ensures that the health care system would not be overburdened with excessive numbers of false positive findings.

In the development of a new test for lung cancer, it is essential to control for the effects of potential confounding variables, because other disorders such as unstable angina [9], breast cancer [8] and diabetes mellitus [11] have also been associated with abnormal breath VOCs. However, a comorbid disorder is unlikely to skew the predictive accuracy of the breath test for two reasons: First, if there are no significant differences between major demographic variables (e.g. age and tobacco smoking) in the disease group and the control group, then the a priori probability of an intercurrent disease is the same in both groups. This is an underlying principle of experimental design in randomized

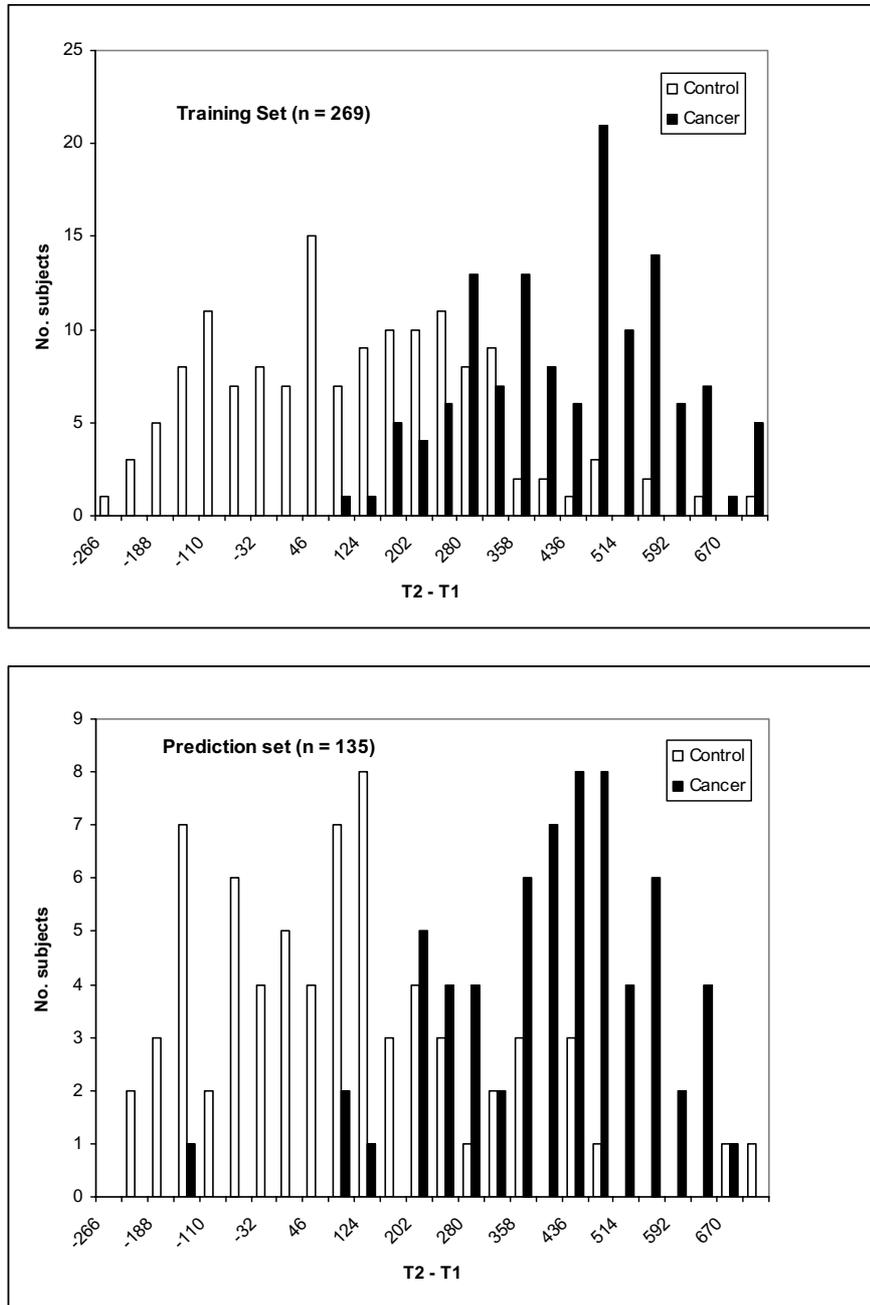


Fig. 8. Typicality scores in cancer patients and controls. Fuzzy logic was employed to construct typicality matrices of selected breath VOC biomarkers of lung cancer in cancer patients and controls respectively. The abundance of each breath VOCs was entered into the typicality matrices in order to generate two numerical values: T_1 , the typicality for no disease, and T_2 , the typicality for primary lung cancer. This histogram demonstrates the distribution of the values of $T_2 - T_1$ in cancer patients and controls, in the training and prediction sets. The mean typicality scores in cancer patients and controls in the prediction set were significantly different (Fig. 2).

clinical trials of new therapeutic interventions, as well as in the evaluation of new diagnostic tests. Second, multivariate modeling has revealed patterns of breath VOCs that are sufficiently distinctive to constitute a

virtual “fingerprint” of a disease. Consequently, the pattern of altered breath VOCs in lung cancer is different from the pattern observed in other disorders such as unstable angina.

We conclude that an intrinsically safe and non-invasive test for breath VOCs combined with fuzzy logic analysis of data provided a rational and accurate predictor of primary lung cancer. A two-minute screening breath test in high risk individuals could potentially identify those with early-stage lung cancer who might benefit most from diagnostic imaging and therapeutic intervention. However, further research will be necessary to determine if detection of lung cancer with this test will be effective in reducing mortality.

Acknowledgements

This research was supported by SBIR award 5R44HL070411-03 from the National Heart Lung and Blood Institute of the National Institutes of Health.

William N Rom was supported by Early Detection Research Network grant UO1 CA086137. Michael Phillips is President and CEO of Menssana Research, Inc. All coauthors had full access to all of the data in the study and they take responsibility for the integrity of the data and the accuracy of the data analysis. We thank James J. Grady DrPH for reviewing the statistical analysis of data.

References

- [1] C. Mountain, Revisions in the International System for Staging Lung Cancer, *Chest* **111** (1997), 1710–1717.
- [2] L. Pauling, A.B. Robinson, R. Teranishi and P. Cary, Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography, *Proc Natl Acad Sci USA* **68** (1971), 2374–2376.
- [3] S.M. Gordon, J.P. Szidon, B.K. Krotoszynski, R.D. Gibbons and H.J. O'Neill, Volatile organic compounds in exhaled air from patients with lung cancer, *Clin Chem* **31** (1985), 1278–1282.
- [4] G.L.J. Preti, J.G. Kostelc, S. Aldinger and R. Daniele, Analysis of lung air from patients with bronchogenic carcinoma and controls using gas chromatography-mass spectrometry, *J Chromatogr* **432** (1988), 1–11.
- [5] M. Phillips, K. Gleeson, J.M. Hughes et al., Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study, *Lancet* **353** (1999), 1930–1933.
- [6] M. Phillips, R. Cataneo, A. Cummin et al., Detection of lung cancer with volatile markers in the breath, *Chest* **123** (2003), 2115–2123.
- [7] M. Phillips, J. Boehmer, R. Cataneo et al., Heart Allograft Rejection: Detection with Breath Alkanes in Low Levels (the HARDBALL study), *The Journal of Heart and Lung Transplantation* **23** (2004), 701–708.
- [8] M. Phillips, R. Cataneo, B. Ditkoff et al., Volatile markers of breast cancer in the breath, *The Breast Journal* **9** (2003), 184–191.
- [9] M. Phillips, R. Cataneo, J. Greenberg, R. Grodman and M. Salazar, Breath markers of oxidative stress in patients with unstable angina, *Heart Disease* **5** (2003), 95–99.
- [10] M. Moretti, M. Phillips, A. Abouzeid, R. Cataneo and J. Greenberg, Increased breath markers of oxidative stress in normal pregnancy and in preeclampsia, *American Journal of Obstetrics and Gynecology* **190** (2004), 1184–1190.
- [11] M. Phillips, R. Cataneo, T. Cheema and J. Greenberg, Increased breath biomarkers of oxidative stress in diabetes mellitus, *Clinica Chimica Acta* **344** (2004), 189–194.
- [12] H. Remmer, T. Hintze, H. Frank and M. Muh-Zange, Cytochrome P-450 oxidation of alkanes originating as scission products during lipid peroxidation, *Xenobiotica* **14** (1984), 207–219.
- [13] S. Zevin and N.L. Benowitz, Drug interactions with tobacco smoking. An update, *Clin Pharmacokinet* **36** (1999), 425–438.
- [14] M. Watanabe, Polymorphic CYP genes and disease predisposition—what have the studies shown so far? *Toxicol Lett* (1998), 102–103: 167–171.
- [15] I. Kapetanovic, S. Rosenfeld and G. Izmirlian, Overview of commonly used bioinformatics methods and their applications, *Ann N Y Acad Sci* **1020** (2004), 10–21.
- [16] C. Pena-Reyes, Evolutionary fuzzy modeling human diagnostic decisions, *Ann N Y Acad Sci* **1020** (2004), 190–211.
- [17] U. Scheipers, H. Ermert, H. Sommerfeld, M. Garcia-Schurmann, T. Senge and S. Philippou, Ultrasonic multifeature tissue characterization for prostate diagnostics, *Ultrasound Med Biol* **29** (2003), 1137–1149.
- [18] L. Shyu, Y. Wu and W. Hu, Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG, *IEEE Trans Biomed Eng* **51** (2004), 1269–1273.
- [19] T. Yin and N. Chiu, A computer-aided diagnosis for distinguishing Tourette's syndrome from chronic tic disorder in children by a fuzzy system with a two-step minimization approach, *IEEE Trans Biomed Eng* **51** (2004), 1286–1295.
- [20] I. Modai, J. Kuperman, I. Goldberg, M. Goldish and S. Mendel, Fuzzy Logic Detection of Medically Serious Suicide Attempt Records in Major Psychiatric Disorders, *J Nerv Ment Dis* **192** (2004), 708–710.
- [21] C.I. Henschke, D.F. Yankelevitz, J.P. Smith, D. Libby, M. Pasmantier, D. McCauley, G. McGuinness, D.P. Naidich, A. Farooqi, M. Vasquez and O.S. Miettinen, CT screening for lung cancer: Assessing a regiment's diagnostic performance, *Clin Imaging* **28** (2004), 317–321.
- [22] M. Phillips, R.N. Cataneo, J. Greenberg, R. Gunawardena, A. Naidu and F. Rahbari-Oskoui, Effect of age on the breath methylated alkane contour, a display of apparent new markers of oxidative stress, *J Lab Clin Med* **136** (2000), 243–249.
- [23] S. Michiels, S. Koscielny and C. Hill, Prediction of cancer outcome with microarrays: a multiple random validation strategy, *Lancet* **365** (2005), 488–492.
- [24] S. Sone, F. Li, Z. Yang et al., Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner, *Br J Cancer* **84** (2001), 25–32.
- [25] D. Ornstein, W. Rayford, V. Fusaro et al., Serum proteomic profiling can discriminate prostate cancer from benign prostates in men with total prostate specific antigen levels between 2.5 and 15.0 ng/ml, *J Urol* **172** (2004), 1302–1305.
- [26] Hr. Bergen, G. Vasmatazis, W. Cliby, K. Johnson, A. Oberg and D. Muddiman, Discovery of ovarian cancer biomarkers in serum using NanoLC electrospray ionization TOF and FT-ICR mass spectrometry, *Dis Markers* **19** (2003–2004), 239–249.

- [27] A. Hoos, H. Hepp, S. Kaul, T. Ahlert, G. Bastert and D. Wallwiener, Telomerase activity correlates with tumor aggressiveness and reflects therapy effect in breast cancer, *Int J Cancer (Pred Oncol)* **79** (1998), 8–12.
- [28] J.C.S. Sharkey, R.J. Behar, R. Perez, J. Otheguy, R. Rabinowitz, J. Steele, C.D.M. Webster, Z. Solc, W. Huff and A. Cantor, Minimally invasive treatment for localized adenocarcinoma of the prostate: review of 1048 patients treated with ultrasound-guided palladium-103 brachytherapy, *J Endourol* **14** (2000), 343–350.
- [29] C.M. Kneepkens, G. Lepage and C.C. Roy, The potential of the hydrocarbon breath test as a measure of lipid peroxidation, *Free Radic Biol Med* **17** (1994), 127–160.
- [30] J.P.G. Schneider, N. Bitterlich, K. Neu, H.G. Velcovsky, H. Morr and N.E.E. Katz, Fuzzy logic-based tumor marker profiles including a new marker tumor M2-PK improved sensitivity to the detection of progression in lung cancer patients, *Anticancer Res* **23** (2003), 899–906.
- [31] S.J.J.J. Swensen, T.E. Hartman, D.E. Midthun, J.A. Sloan, A.M. Sykes, G.L. Aughenbaugh and C. MA, Lung cancer screening with CT: Mayo Clinic experience, *Radiology* **226** (2003), 756–761.
- [32] C.I. Henschke, D.I. McCauley, D.F. Yankelevitz et al., Early Lung Cancer Action Project: overall design and findings from baseline screening, *Lancet* **354** (1999), 99–105.
- [33] S. Baker, B. Kramer, M. McIntosh, B. Patterson, Y. Shyr and S. Skates, Evaluating markers for the early detection of cancer: overview of study designs and methods, *Clinical Trials* **3** (2006), 43–56.