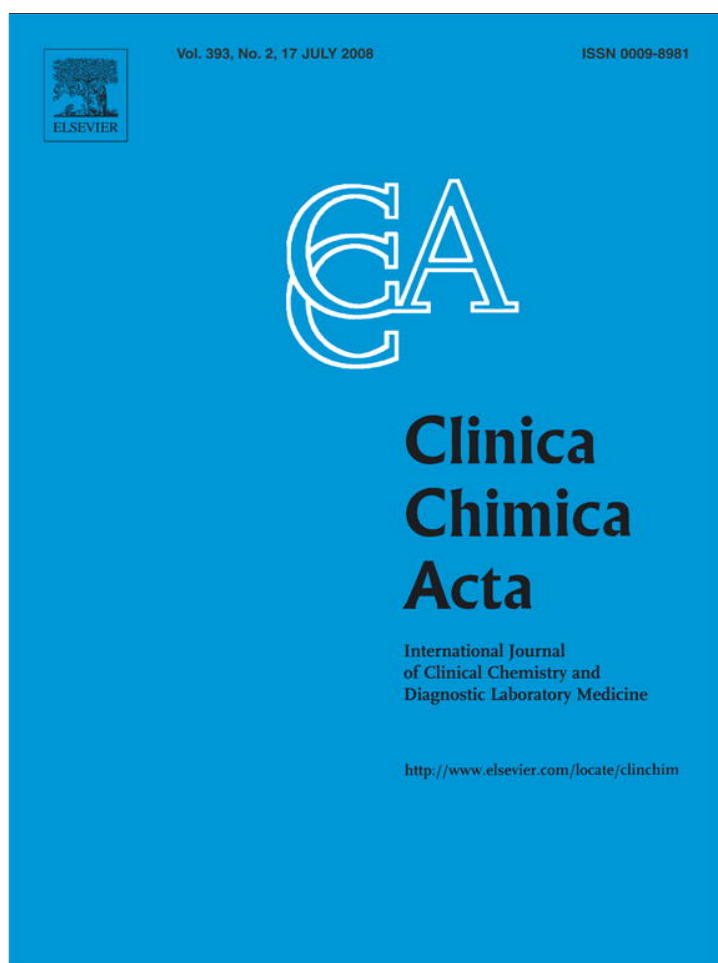


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

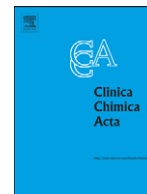
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: [www.elsevier.com/locate/clinchim](http://www.elsevier.com/locate/clinchim)

## Detection of lung cancer using weighted digital analysis of breath biomarkers

Michael Phillips<sup>a,b,\*</sup>, Nasser Altorki<sup>c</sup>, John H.M. Austin<sup>d</sup>, Robert B. Cameron<sup>f</sup>, Renee N. Cataneo<sup>a</sup>, Robert Kloss<sup>g</sup>, Roger A. Maxfield<sup>e</sup>, Muhammad I. Munawar<sup>a,1</sup>, Harvey I. Pass<sup>h</sup>, Asif Rashid<sup>a,2</sup>, William N. Rom<sup>i</sup>, Peter Schmitt<sup>j</sup>, James Wai<sup>a</sup>

<sup>a</sup> Menssana Research Inc, Fort Lee, NJ 07024-6510, United States

<sup>b</sup> Department of Medicine, New York Medical College, Valhalla, NY, United States

<sup>c</sup> Division of Thoracic Surgery, Department Of Cardiothoracic Surgery, Weill Medical College of Cornell University, New York, NY 10021, United States

<sup>d</sup> Department of Radiology, Columbia University Medical Center, New York, NY 10032, United States

<sup>e</sup> Department of Medicine, Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University Medical Center, New York, NY 10032, United States

<sup>f</sup> Divisions of Thoracic Surgery and Surgical Oncology, UCLA School of Medicine and West Los Angeles VA Medical Center, Los Angeles, CA 90095-7313, United States

<sup>g</sup> Danbury Hospital, Danbury, CT 06810, United States

<sup>h</sup> Division of Thoracic Surgery and Thoracic Oncology, Department of Cardiothoracic Surgery, New York University Medical Center, New York, NY 10016, United States

<sup>i</sup> Division of Pulmonary and Critical Care Medicine, New York University Medical Center, New York, NY 10016, United States

<sup>j</sup> Schmitt & Associates, 211 Warren St, Newark, NJ 07103, United States

### ARTICLE INFO

#### Article history:

Received 28 July 2007

Received in revised form 15 February 2008

Accepted 26 February 2008

Available online 3 March 2008

#### Keywords:

Breath test

Lung cancer

Diagnosis

Multivariate analysis

Biomarkers

### ABSTRACT

**Background:** A combination of biomarkers in a multivariate model may predict disease with greater accuracy than a single biomarker employed alone. We developed a non-linear method of multivariate analysis, weighted digital analysis (WDA), and evaluated its ability to predict lung cancer employing volatile biomarkers in the breath.

**Methods:** WDA generates a discriminant function to predict membership in disease vs no disease groups by determining weight, a cutoff value, and a sign for each predictor variable employed in the model. The weight of each predictor variable was the area under the curve (AUC) of the receiver operating characteristic (ROC) curve minus a fixed offset of 0.55, where the AUC was obtained by employing that predictor variable alone, as the sole marker of disease. The sign ( $\pm$ ) was used to invert the predictor variable if a lower value indicated a higher probability of disease. When employed to predict the presence of a disease in a particular patient, the discriminant function was determined as the sum of the weights of all predictor variables that exceeded their cutoff values. The algorithm that generates the discriminant function is deterministic because parameters are calculated from each individual predictor variable without any optimization or adjustment. We employed WDA to re-evaluate data from a recent study of breath biomarkers of lung cancer, comprising the volatile organic compounds (VOCs) in the alveolar breath of 193 subjects with primary lung cancer and 211 controls with a negative chest CT.

**Results:** The WDA discriminant function accurately identified patients with lung cancer in a model employing 30 breath VOCs (ROC curve AUC=0.90; sensitivity=84.5%, specificity=81.0%). These results were superior to multilinear regression analysis of the same data set (AUC=0.74, sensitivity=68.4, specificity=73.5%). WDA test accuracy did not vary appreciably with TNM (tumor, node, metastasis) stage of disease, and results were not affected by tobacco smoking (ROC curve AUC=0.92 in current smokers, 0.90 in former smokers). WDA was a robust predictor of lung cancer: random removal of 1/3 of the VOCs did not reduce the AUC of the ROC curve by >10% (99.7% CI).

**Conclusions:** A test employing WDA of breath VOCs predicted lung cancer with accuracy similar to chest computed tomography. The algorithm identified dependencies that were not apparent with traditional linear methods. WDA appears to provide a useful new technique for non-linear multivariate analysis of data.

© 2008 Elsevier B.V. All rights reserved.

### 1. Introduction

Most diagnostic classifications are binary e.g., dead or alive, disease or no disease, cancer or no cancer. The physician's task is to correctly assign a patient to one group or the other by applying generally accepted criteria of membership. This presents little difficulty if the difference between the two groups is defined by a single criterion that

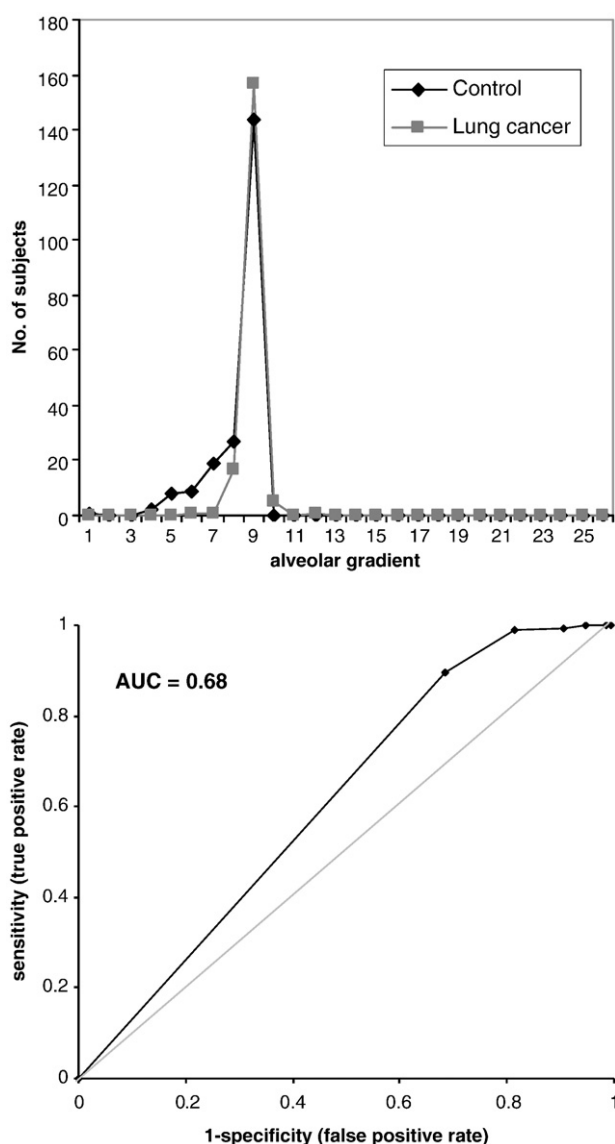
\* Corresponding author. Menssana Research, Inc., 1 Horizon Road, Suite 1415, Fort Lee, NJ 07024, United States. Tel./fax: +1 201 886 7004.

E-mail address: [mphillips@menssana-research.com](mailto:mphillips@menssana-research.com) (M. Phillips).

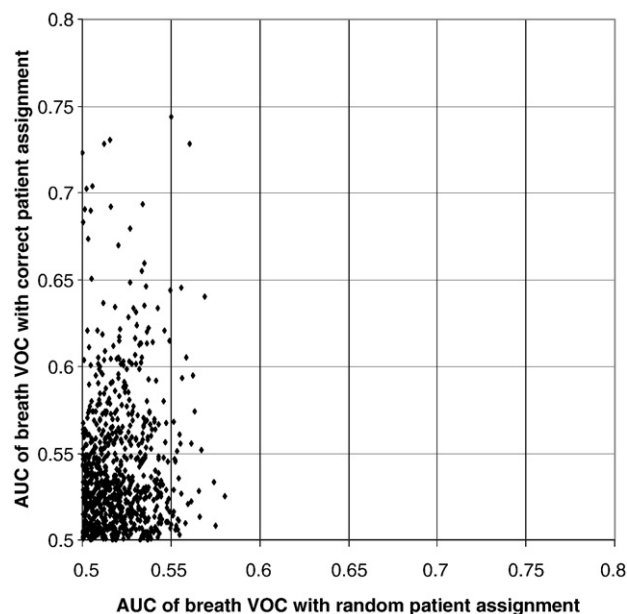
<sup>1</sup> Present address: Department of Psychiatry & Human Behavior, University of Mississippi Medical Center, Jackson, MS 39216, United States.

<sup>2</sup> Present address: Saint Vincent Catholic Medical Centers, Staten Island Campus, Staten Island, NY 10310, United States.

can be readily estimated e.g., hypoglycemia can be distinguished from normoglycemia by measuring the blood glucose concentration, and then determining whether it is above or below a designated cutoff value. However, clinical diagnosis is usually more difficult because assignment to one group or the other generally requires a combination of several different criteria. For example, a patient with acute streptococcal pneumonia may present with several different symptoms and signs, including fever, chills, productive cough, herpes labialis, vocal fremitus, and localized crackles in the chest. As an additional complication, not all of these features are required for the diagnosis, and some are more important than others. An experienced physician takes these difficulties into account by intuitively assigning a different diagnostic weight to each finding. For example, fever has a comparatively low diagnostic weight because it may be absent in elderly or immune suppressed patients, but the diagnostic weight of chest



**Fig. 1.** Accuracy of a single breath VOC employed as a biomarker of lung cancer. Upper panel: distribution of alveolar gradients of isopropyl alcohol in lung cancer patients and in controls. The sensitivity and the specificity of this VOC as a biomarker of lung cancer varies with the cutoff value at different points along the x-axis. Lower panel: the receiver operating characteristic (ROC) curve derived from the sensitivity and the specificity observed at different cutoff values along the x-axis. For isopropyl alcohol in breath, the AUC of the ROC curve was 0.68, indicating that it was a modestly accurate biomarker of lung cancer when employed alone.



**Fig. 2.** Effect of random patient assignment on predictive accuracy. All breath VOCs were evaluated as biomarkers of lung cancer employing the method shown in Fig. 1. The AUC of each ROC curve is displayed employing the correct cancer/control assignment (y-axis) vs random assignment to the cancer or the control group (x-axis). This figure demonstrates the difference between the 2 distributions: when the diagnosis was randomly assigned, no VOC ROC curve had an individual AUC of  $\geq 0.6$ . However, when the diagnosis was correctly assigned, 69 VOCs had a ROC curve AUC  $> 0.6$ , and these VOCs were selected as the best biomarkers of lung cancer.

crackles is much higher. This process of intuitively weighing all of these relative values and then incorporating them into a binary prediction is often described as “diagnostic clinical judgment”.

Since the exercise of diagnostic clinical judgment is an intuitive process, its outcome must necessarily vary with an individual's skill and experience. As a consequence, physicians have sought to improve the accuracy and reproducibility of clinical judgment by employing formal algorithms to assign patients to appropriate diagnostic groups. As early as 1944, Jones proposed an algorithm for the diagnosis of rheumatic fever that employed a combination of “high weight” major criteria (including carditis and polyarthritis) and “low weight” minor criteria (including fever and arthralgia). As evidence of its clinical value, a modified version of this algorithm is still in clinical use  $> 60$  y later [1]. The accuracy and reproducibility of diagnostic laboratory tests can also be improved in the same fashion. A predictive algorithm employing the relative diagnostic weights of two or more biomarkers of lung cancer in combination can predict disease with greater accuracy than a single biomarker employed alone [2].

Non-linear multivariate statistical analysis provides a useful tool for determining the relative weights of clinical markers of disease and incorporating them into new diagnostic algorithms. We have previously reported that biomarkers in the breath predict lung cancer [3,4], breast cancer [5], pulmonary tuberculosis [6], and heart transplant rejection [7]. All of these tests assigned a relative weight to a number of different biomarkers – volatile organic compounds (VOCs) in the breath – and incorporated them into a predictive algorithm with a binary outcome i.e. disease or no disease. A non-linear multivariate model employing fuzzy logic predicted lung cancer with greater accuracy than multilinear analysis [8].

We report here a new method for non-linear modeling. Weighted digital analysis (WDA) determines the relative weights of a set of VOCs as biomarkers of disease, and incorporates them into an algorithm to predict the presence or absence of disease. We present evidence for

the effectiveness of WDA in a reanalysis of data obtained from a previous study of breath VOC biomarkers of lung cancer.

2. Methods

2.1. Clinical study

This has been previously reported [8]. In summary, breath samples were collected from 404 patients: 193 with untreated primary lung cancer and 211 controls with no evidence of cancer on chest CT. VOCs in alveolar breath and ambient air were analyzed by gas chromatography and mass spectroscopy. The data set comprised the breath VOCs in patients with untreated primary lung cancer and the controls.

2.2. Determination of alveolar gradients

The alveolar gradient is the difference between the abundance of a VOC in breath and air, and the method for its determination has been described [9].

2.3. General principles of WDA

WDA is a mathematical method for developing a diagnostic algorithm that generates a discriminatory function score. The value of this score predicts membership in 1 of 2 groups e.g., disease or no disease. Every diagnostic variable (e.g., the numerical value of a laboratory test result) that is employed in the algorithm has three parameters.

- a. *Sign*. This may be positive (+1) or negative (−1). If the sign is positive, a higher value of the diagnostic variable indicates that disease is more likely. Conversely, if its sign is negative, a lower value of the variable indicates that disease is more likely.
- b. *Weight*. This value indicates the relative contribution of each diagnostic variable to the discriminatory function, i.e. the higher its weight, the greater is its relative importance as a predictor of disease.
- c. *Cutoff value*. This determines whether or not a diagnostic variable contributes to the discriminatory function score. The contribution of a diagnostic variable (weight) is

added to the discriminatory function score only if that value exceeds the cutoff value.

For each variable, the value of (gradient X sign) is determined, and if it exceeds the (cutoff X sign) value, the weight is added to the discriminatory function score. The discriminatory function score (df) for a given patient (subscript *i*) is calculated as:

$$df_i = \sum_c y_{ic} \quad \text{with } y_{ic} = \begin{cases} \text{Weight}_c & \text{if } x_{ic} * \text{Sign}_c > \text{Cutoff}_{c*} \text{Sign}_c \\ 0 & \text{otherwise} \end{cases}$$

The subscript *c* incorporates all contributing VOCs.

The alveolar gradient of each VOC was employed to create a receiver operating characteristic (ROC) curve, demonstrating its ability to distinguish cancer patients from controls. As an example, the ROC curve for a breath VOC tentatively identified as isopropyl alcohol is shown in Fig. 1. From this ROC curve, the area under curve (AUC), sign (−1 if AUC is <0.5, +1 if AUC is >=0.5), and cutoff value *C* of the alveolar gradient where [sensitivity + (1 − specificity)] is a maximum were determined. This enables the determination of the cutoff (*C*),  $\text{weight} = (\text{Abs}(\text{AUC} - 0.5) + 0.5) - 0.55$ , and  $\text{sign} = -1$  if AUC is <0.5; +1 if AUC is >=0.5. The WDA algorithm may be readily employed in a computerized spreadsheet program. It is currently implemented as a macro in Microsoft Excel.

2.4. Selection of VOCs for inclusion in WDA

ROC curves were determined in this fashion for all breath VOCs, and only those VOCs with weight >0.6 were selected for further analysis. Breath VOCs were edited to combine duplicates i.e. ROC curves were constructed from the discriminatory function generated by employing all of the predictor variables in the model.

2.5. Robustness of the model

We defined robustness as the number of breath VOCs that may be lost from the model without incurring a significant deterioration in its predictive accuracy. In this context, a “lost” VOC is either absent or else it is generating random results. We

Table 1  
Major VOC identifiers of primary lung cancer in breath

	Breath VOC	CAS #	NIST #
1	Isopropyl alcohol	67-63-0	229015
2	4-Penten-2-ol	625-31-0	235505
3	Ethane, 1,1,2-trichloro-1,2,2-trifluoro-	76-13-1	233813
4	Propane, 2-methoxy-2-methyl-	1634-04-4	229277
5	1-Propene, 1-(methylthio)-, (E)-	42848-06-6	26402
6	2,3-Hexanedione	3848-24-6	291460
7	5,5-Dimethyl-1,3-hexadiene	1515-79-3	113453
8	3-Hexanone, 2-methyl-	7379-12-6	231728
9	1H-Indene, 2,3-dihydro-4-methyl-	824-22-6	2991
10	Camphor	21368-68-3	73611
11	Bicyclo[2.2.1]heptan-2-one, 1,7,7-trimethyl-, (1S)-	464-48-2	114690
12	3-Cyclohexene-1-methanol, $\alpha$ , $\alpha$ -4-trimethyl-	98-55-5	231634
13	<i>p</i> -menth-1-en-8-ol	N/A	151924
14	5-Isopropenyl-2-methyl-7-oxabicyclo[4.1.0]heptan-2-ol	N/A	185009
15	$\alpha$ Isomethyl ionone	127-51-5	196736
16	2,2,7,7-Tetramethyltricyclo[6.2.1.0(1,6)]undec-4-en-3-one	N/A	189499
17	2,2,4-Trimethyl-1,3-pentanediol diisobutyrate	6846-50-0	151177
18	Benzoic acid, 4-ethoxy-, ethyl ester	23676-09-7	107721
19	Bicyclo[3.2.2]nonane-1,5-dicarboxylic acid, 5-ethyl ester	24238-73-1	269408
20	Pentanoic acid, 2,2,4-trimethyl-3-carboxyisopropyl, isobutyl ester	N/A	140775
21	Propanoic acid, 2-methyl-, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester	74381-40-1	59556
22	1,2,4,5-Tetroxane, 3,3,6,6-tetraphenyl-	16204-36-7	11836
23	Benzophenone	119-61-9	118652
24	2,5-Cyclohexadien-1-one, 2,6-bis(1,1-dimethylethyl)-4-ethylidene-	6738-27-8	215417
25	Furan, 2-[(2-ethoxy-3,4-dimethyl-2-cyclohexen-1-ylidene)methyl]-	55162-49-7	47619
26	Benzene, 1,1-(1,2-cyclobutanediyl)bis-, <i>cis</i> -	7694-30-6	62825
27	Benzene, 1,1-[1-(ethylthio)propylidene]bis-	53699-80-2	149972
28	Anthracene, 1,2,3,4-tetrahydro-9-propyl-	101580-33-0	155542
29	9,10-Anthracenediol, 2-ethyl-	839-73-6	153923
30	Benzene, 1,1-ethylidenebis[4-ethyl]-	10224-91-6	11431

WDA selected these 30 breath VOCs as candidate biomarkers of primary lung cancer because the AUC of each ROC curve exceeded 0.6. VOCs are ranked by their chromatographic retention times. CAS and NIST numbers are shown, where available. N/A = not available. In a previous report [8], fuzzy logic selected a different set of lung cancer biomarkers from the same data set, possibly reflecting that fuzzy logic and WDA are fundamentally different techniques of multivariate analysis. Two sets of duplicate VOCs were identified in this list (10, 11 and 12, 13, highlighted). This was a consequence of the mass spectra library assigning different synonyms (with different CAS and NIST numbers) to the same VOC on different occasions. This did not affect the outcome of multivariate analysis with WDA.

employed the following algorithm to evaluate robustness: select the candidate VOCs in the model, rank these VOCs by AUC of the ROC curve:

1. Determine their cumulative AUC.
2. Randomly remove one VOC at a time until the cumulative AUC falls below 90% of the original value.
3. Repeat this step  $n$  (e.g., 30) times.
4. The average number of removed VOCs is termed the robustness for 10% degradation.

### 2.6. Detection of lung cancer

We employed WDA to analyze alveolar gradients of breath VOCs in the entire data set of 404 subjects (untreated primary lung cancer and cancer-free controls) employing the method described above. In addition, we cross-validated algorithms in randomly split subset groups [10]. Subjects were randomly assigned to a training set or to a prediction set in a 2:1 ratio. WDA was performed employing multiple ( $n=20$ ) randomly selected unique training and prediction sets.

## 3. Results

### 3.1. Candidate biomarkers of lung cancer

Fig. 1 displays the accuracy of a typical single breath VOC employed as a biomarker of lung cancer. Note that this VOC alone did not significantly distinguish between cancer patients and the control group; it required the combined results of several VOCs to generate a discriminant function with a significant AUC value. Fig. 2 displays the effect of random patient assignment on predictive accuracy of individual breath VOCs, and how these results were employed to identify the optimal candidate biomarkers of lung cancer, with AUC of ROC curve  $>0.6$ . This cutoff value was employed because no VOC had an AUC  $>0.6$  when the diagnosis was randomly assigned (Table 1).

### 3.2. WDA discriminatory function in all cases of lung cancer

Fig. 3 displays WDA discriminatory function values in controls and lung cancer patients. The area under curve (AUC) of the resulting receiver operating characteristic (ROC) curve was 0.90. Employing multilinear regression of the same data set, AUC of ROC curve = 0.74.

### 3.3. WDA discriminatory function in lung cancer stratified by TNM (tumor, node, metastasis) stage

Fig. 3 displays mean discriminatory function values in controls and lung cancer patients stratified by TNM stages 1 to 4. Fig. 4 displays the ROC curves obtained from these data. Test accuracy did not vary appreciably with TNM stage of disease.

### 3.4. WDA discriminatory function in lung cancer stratified by tobacco smoking

Fig. 5 displays the ROC curves obtained when WDA data in Fig. 3 was stratified according to whether subjects were current smokers (AUC=0.92) or former smokers (AUC=0.90). By inspection, the similarity of the two ROC curves demonstrated that the WDA discriminatory function was not affected by whether subjects were current smokers or ex-smokers.

### 3.5. Effect of the number of VOCs in model on WDA discriminatory function

Fig. 6 displays the effect of the number of VOCs employed in the model on the AUC of the ROC curve for all patients with lung cancer. VOC biomarkers of lung cancer shown in Table 1 were added to the model one by one, commencing with the highest weight VOC. The WDA discriminatory function required only ten VOCs in order to identify lung cancer with near maximal accuracy. However, more VOCs were added to the algorithm in order to enhance the robustness of the analysis.

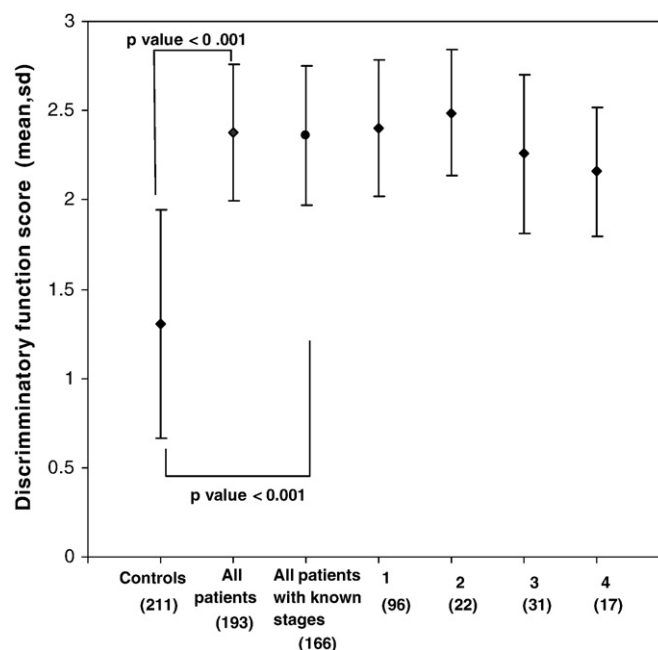
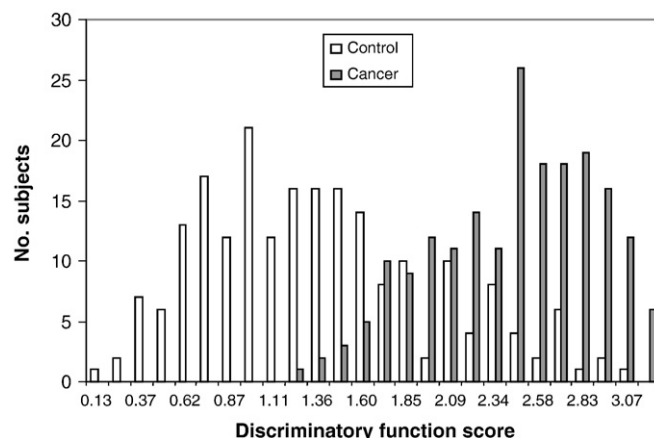


Fig. 3. WDA discriminatory function scores in lung cancer and controls. Upper panel: This histogram displays the distribution of discriminatory function scores in the 2 groups. Lower panel: Mean discriminatory function scores in controls and patients with lung cancer stratified by TNM stage of disease. TNM staging information was available for 166/193 patients with lung cancer. Mean discriminatory function scores were 2.36 (SD=0.47) in all stages of lung cancer, and 1.30 (SD=0.64) in controls ( $p < 10^{-4}$ , 2 tailed  $t$ -test). (Since the overall AUC of the total set is high (around 0.9) it is to be expected that the AUC of any of the subsets stratified by TNM stage will have a similarly high value.)

### 3.6. Cross validation in random split subsets

Fig. 7 displays the training and validation ROC curves (mean of 20 random splits). The same 30 VOCs employed in the final model were also employed in these split data sets; however the cutoff points, signs, and weight, were adjusted for each split based on the results in their respective training sets.

### 3.7. Robustness of the model

Fig. 8 displays variation in the robustness with the number of VOCs in the model, as well as “robustness –3 Sigma”, which indicates the number of VOCs that can be lost with 99.7% probability the AUC will not degrade by  $>10\%$ . If, for example, the model employed 30 VOCs, a third of these VOCs could be lost without reducing the accuracy of the breath test by  $>10\%$  at 99.7% CI.

### 3.8. Effect of random assignment of diagnosis

Fig. 7 (lower panel) displays the effect of random reversal of assignment of patient diagnosis, prior to determination of the WDA discriminatory function scores. The accuracy of the WDA model progressively deteriorated with the declining integrity of the breath VOC data, supporting the conclusion that the undegraded WDA model identified lung cancer by extracting a signal of disease from the breath VOC data.

## 4. Discussion

A test employing WDA of a combination of breath VOCs accurately identified patients with lung cancer. The accuracy of the breath test may be directly compared to that of chest CT. In a large population screening study, chest CT detected lung cancer with 55% sensitivity and 95% specificity [11]. As the ROC curve in Fig. 4 demonstrates, at the point where the breath test sensitivity was 55%, its specificity was approximately 93% i.e. close to the same as chest CT.

Table 1 lists the breath VOCs identified as candidate biomarkers of primary lung cancer. Although we observed a strong statistical association between lung cancer and a set of apparent VOC biomarkers in the breath, the biological mechanism linking lung

cancer with these breath VOCs has not yet been identified with certainty. Tumor markers are conventionally regarded as downstream products that are manufactured in cancer cells and discharged into the blood. Examples include CA125 in ovarian cancer, PSA in prostate cancer and CEA in ovarian and breast cancers. However, we have previously described four important points of difference between typical downstream tumor markers and breath VOC biomarkers of lung cancer [8].

### 4.1. Biological significance and variation

Few of the breath VOCs associated with lung cancer have known biological significance in lung disease. Also, the set of breath VOCs associated with lung cancer has been found to vary from study to study, and also within an individual study when different techniques of multivariate analysis were employed.

### 4.2. Effect of tumor mass

Serum levels of a tumor marker may increase as a tumor grows larger [13]. However, tumor mass did not affect the abundance of breath VOC biomarkers of lung cancer; these remained relatively constant, as shown by the similarity of ROC curves in TNM stages 1 through 4 (Fig. 4).

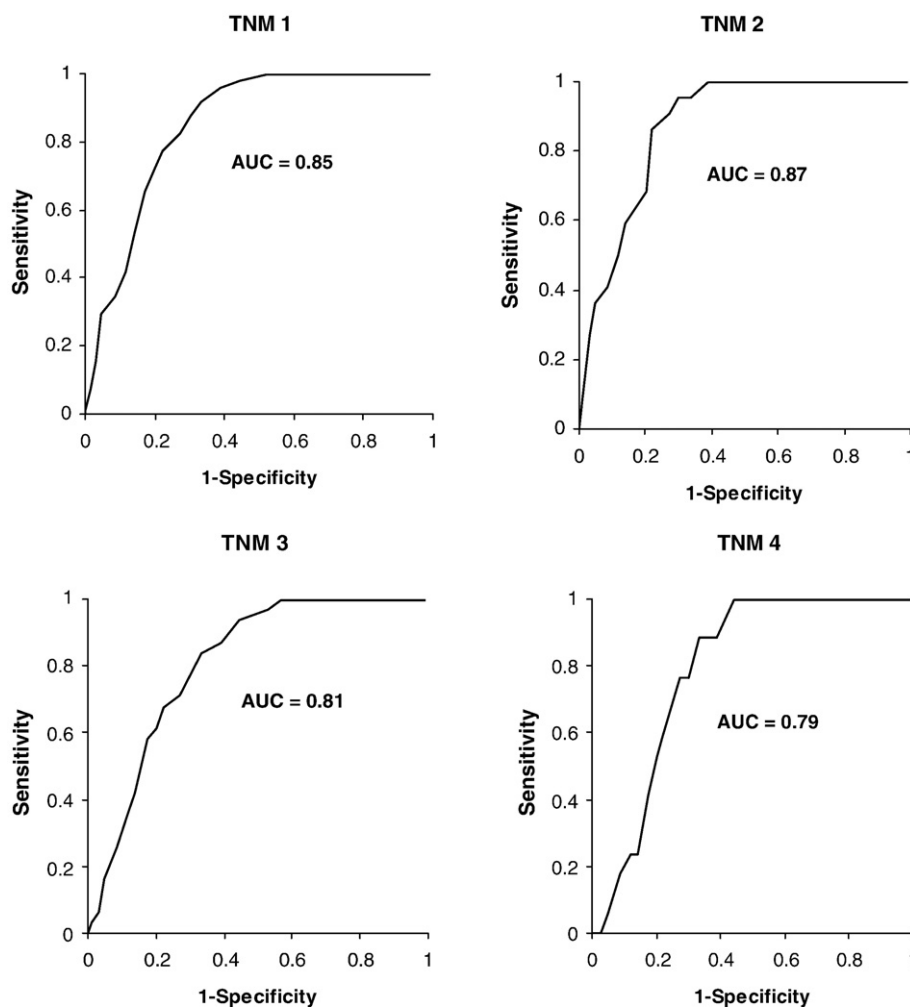
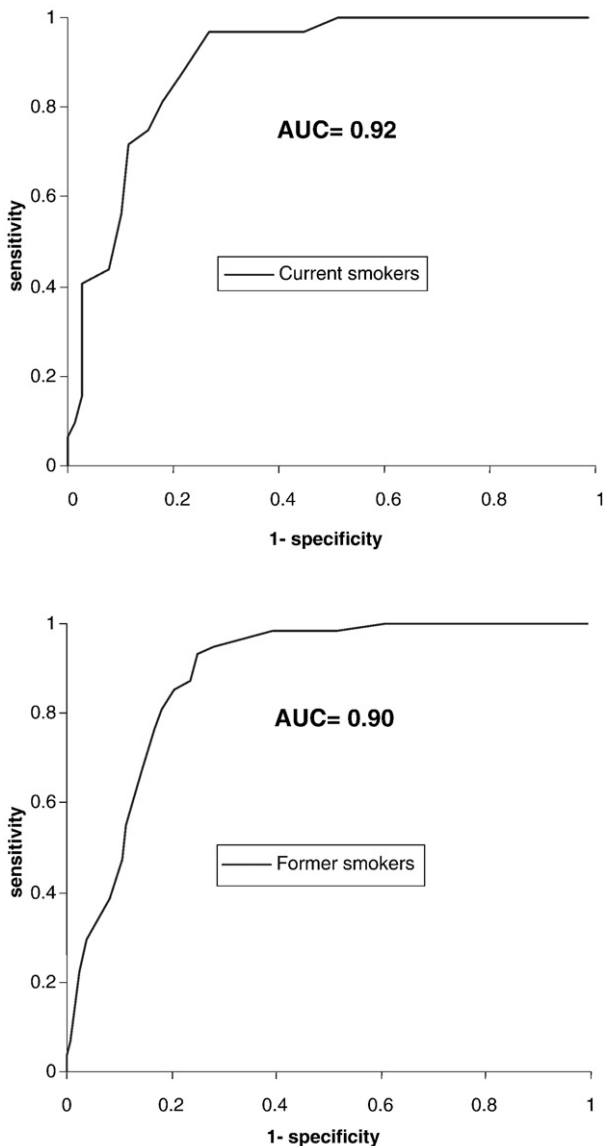


Fig. 4. Breath biomarkers of lung cancer stratified by TNM stage: These figures display the ROC curves obtained by stratifying the WDA data in Fig. 4 according to the TNM stage of lung cancer. The AUC was high in TNM1 lung cancer, and a similar performance was maintained at all other stages. (Since the overall AUC of the total set is high (around 0.9) it is to be expected that the AUC of any of the subsets stratified by TNM stage will have a similarly high value.)



**Fig. 5.** Breath biomarkers of lung cancer stratified by tobacco smoking: These figures display the ROC curves obtained by stratifying the WDA data in Fig. 4 according to whether subjects were current smokers (upper panel; AUC=0.92 or former smokers (lower panel; AUC=0.90), demonstrating that the WDA discriminatory function scores were not skewed by current smoking or a history of smoking.

4.3. Effect of surgery

Ablation of the prostate reduces serum levels of PSA [14], and concentrations of serum tumor marker are generally reduced by excision of the cancer. However, we previously reported that the outcome of the breath test was unchanged in most patients with lung cancer following thoracotomy with resection of the tumor.

4.4. Abundance

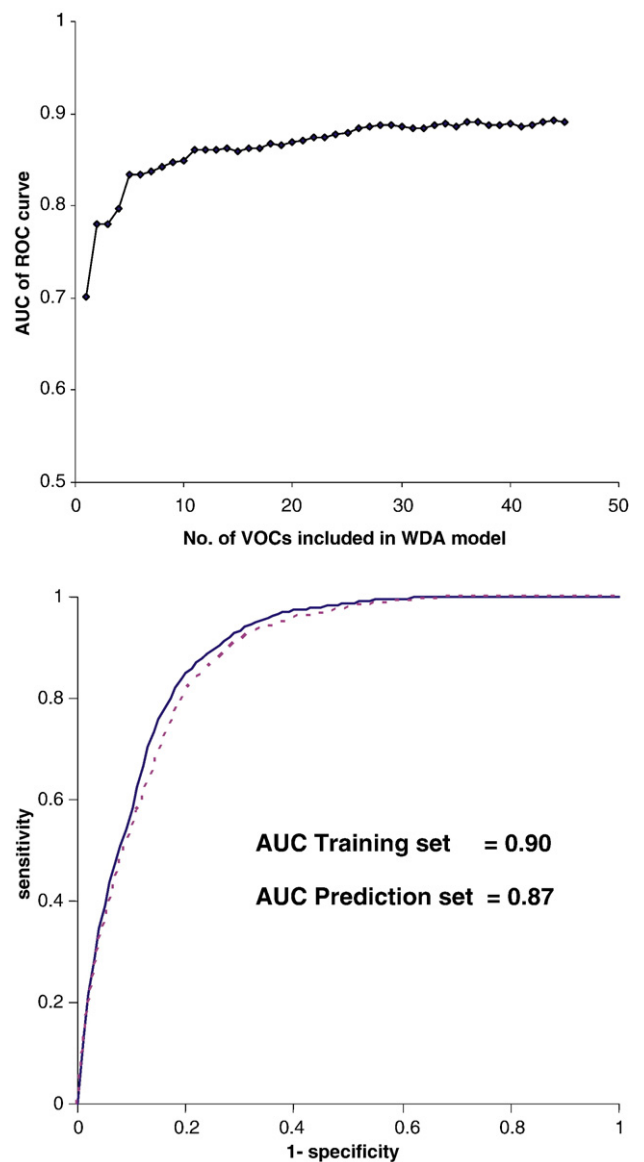
Serum tumor markers are consistently increased in patients with cancer; however, we observed a combination of decreased as well as increased abundance of breath VOC biomarkers in lung cancer.

For these reasons, we concluded that the downstream model of tumor marker production did not provide a satisfactory explanation of the observed breath VOC biomarkers of lung cancer. We therefore proposed an alternative biological mechanism: an upstream model hypothesis, in which the pathophysiologic process that results in lung

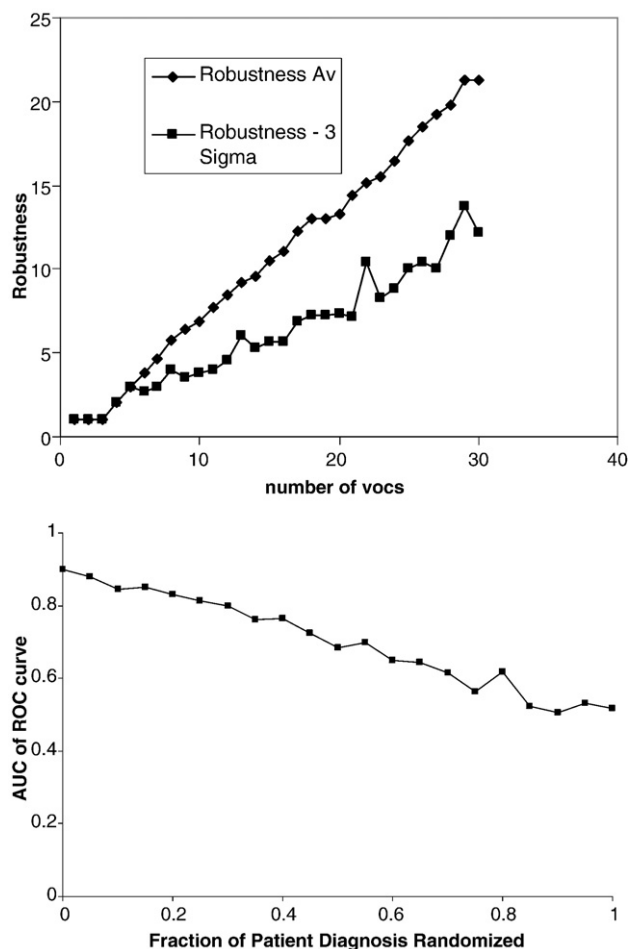
cancer may also modulate the abundance of VOCs in breath, so that carcinogenesis and altered breath VOCs are 2 concurrent but independent phenomena. Fig. 8 displays a pathophysiologic hypothesis in which activation of lethal cytochrome p450 mixed oxidases may lead to lung cancer while independently altering the catabolism of VOCs. This model provides a rational explanation for the observed points of difference described above:

4.5. Biological significance and variation

More than 3000 different VOCs have been observed in normal human breath [9], all of them with low molecular weights (<600), unlike protein serum tumor markers which have molecular weights of several kilodaltons [15,16]. Induced cytochrome p450



**Fig. 6.** Effect of the number of VOCs on model performance, and cross validation in random split subsets. Upper panel: The accuracy of the breath test varied with the number of VOCs employed in the model. VOC biomarkers of lung cancer were added to the model one by one, commencing with the highest weight VOC. This figure demonstrates that the breath test identified lung cancer with near maximal accuracy with only 10 VOCs. Lower panel: Mean ROC curves of breath test results employing the same 30 VOCs in 20 random split data sets into a training set and a test set in a 2:1 ratio. The cutoff points, signs, and weight, were adjusted for each split based on the results in their respective training sets.



**Fig. 7.** Robustness of the WDA model and effect of random diagnosis assignment: Upper panel: This figure displays “robustness” vs the number of VOCs included in the analysis. Robustness is defined by the number of VOCs that can be removed on average without degrading the AUC of the ROC curve by >10%. The value is derived by removing randomly selected VOCs from the analysis until the AUC drops by 10%. The line “Robustness – 3 Sigma” indicates the number of VOCs that can be lost so that with 99.7% probability the AUC will not degrade by >10%. When the WDA analysis included 30 VOCs (arrow), the value of “Robustness – 3 Sigma” was approximately 10. This indicates that a third of the VOCs could be lost from the model without reducing its accuracy by >10% at the 99.7% confidence level. In this context “lost” means that the VOCs were not present in a patient’s breath sample or in the room air. Lower panel: This figure displays the effect of the fraction of patients randomized on the AUC of the ROC curve. Assignment of patients to the cancer or the control group was randomized prior to determination of the WDA discriminatory function scores. The accuracy of the WDA model progressively deteriorated with the addition of random classifiers: the AUC of the ROC curve degraded approximately 4% for every 10% of random classifier changes. This supports the conclusion that the undegraded WDA model extracted a lung cancer signal from breath VOCs, because the accuracy of detection fell with the declining integrity of the signal.

mixed oxidase activity could potentially modulate the catabolism of many of these breath VOCs, and thereby account for the large and diverse sets of candidate breath biomarkers associated with lung cancer. Cytochrome p450 enzymes catabolize most of the VOCs listed in Table 1, including isopropyl alcohol [17], hexanediene [18], camphor [19], benzophenone [20] and derivatives of tetraoxane [21], benzene [22], benzoic acid [23], furan [24] and ionone [25] (this list is not exhaustive). The resulting diversity of candidate biomarkers constitutes a major strength of breath testing for lung cancer, since it ensures redundancy and robustness of the predictive algorithm. As Fig. 7 demonstrates, several of these VOCs can be removed from the model without any significant degradation of the accuracy of the test as a predictor of lung cancer.

#### 4.6. Effects of tumor mass and surgery

The model shown in Fig. 8 predicts that tumor mass will have no effect on breath biomarkers of lung cancer, since induction of cytochrome p450 activity would alter the composition of breath VOCs independently of the growth of lung cancer cells. Similarly, resection of the lung cancer should have no effect on the breath signal.

#### 4.7. Effects on breath VOC abundance

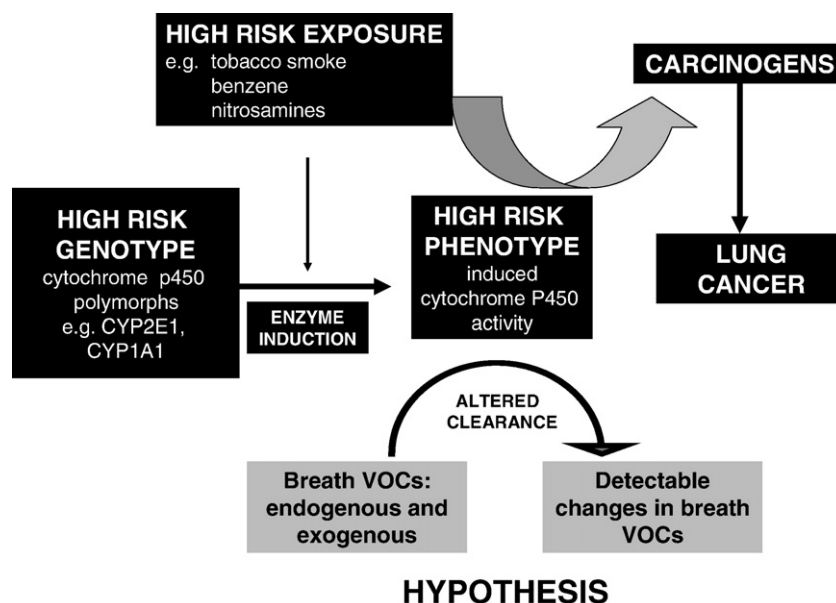
This model predicts a combination of decreased as well as increased abundance of breath VOC, since VOC precursors will be depleted by catabolism, while their metabolites will be increased.

Table 2 lists the advantages of WDA compared to traditional multilinear analysis, and newer methods of multivariate data analysis such as fuzzy logic, pattern recognition, or neural networks. WDA identified a set of VOC biomarkers of lung cancer that were similar, though not identical, to those identified in a previous study employing fuzzy logic analysis of data [8]. Other results were also similar to those previously reported [8]. The accuracy of the breath test was not affected by the TNM stage of lung cancer, nor by current or former tobacco smoking.

There are risks as well as benefits to adding more variables to a multivariate predictive algorithm. While the main benefit is improved accuracy of the algorithm, the risk is that some of this improvement may be illusory, because the inclusion of variables with poorer correlations can degrade predictive accuracy. Consequently, the number of variables in an algorithm is usually a compromise between two conflicting demands: there must be a sufficient number to ensure accuracy, yet not so many as to introduce spurious results. We identified 69 candidate VOC biomarkers of lung cancer in the breath (Fig. 2), and then ranked them according to the AUC of their ROC curves. We selected the top 30 VOCs for inclusion in the predictive algorithm based solely on their AUC values, without any knowledge or consideration of their potential biological significance. This approach yielded three main advantages: first, the algorithm was highly sensitive and specific for lung cancer – the AUC of the ROC curve was close to 0.9 (Fig. 6, upper panel). Second, there was no evidence of spurious results caused by poorly correlated variables because the cross-validated ROC curve exhibited virtually no degradation (Fig. 6, lower panel). Third, the WDA model was highly robust. When one third of the VOCs were randomly removed from a 30 VOC model, the accuracy of the breath test was degraded by less than 10% at the 99.7% confidence level (Fig. 7, upper panel). This provides evidence for high redundancy in breath testing for lung cancer: Even when part of the information in the breath VOC signal was unavailable, the WDA model delivered highly accurate information about the presence or absence of disease.

The weight of each predictor variable was the area under the curve (AUC) of its receiver operating characteristic (ROC) curve. We employed only VOCs with  $AUC \geq 0.6$ , and subtracted a fixed offset of 0.55 from these values in order to increase the relative differences between them. For example, if VOC1  $AUC = 0.6$  and VOC2  $AUC = 0.65$ , then VOC2 has approximately 10% higher weight without subtraction. However, with subtraction of 0.55, the relative AUC of VOC2 to VOC1 is doubled. Most of the AUCs in this data set fell between 0.6 and 0.7, so that their relative differences were markedly increased by this procedure. In practice, employment of this arbitrary offset was justified by the improvement in the resulting discriminant function.

Lung cancer causes more deaths than any other malignancy in the U.S. [26]. Since annual screening with chest computed tomography (CT) can detect early stage lung cancer that is likely to be curable [27], there is hope that early detection could increase 5-y survival. However, chest CT screening is costly and the hazards of associated radiation may outweigh its potential benefits [28]. These concerns have led researchers to seek biomarkers of lung cancer that could provide



**Fig. 8.** Hypothetical basis of the breath test for lung cancer: Lung cancer may result from the interaction of hereditary and environmental factors. A person's genotype may include a variety of cytochrome p450 mixed oxidases, some of which are activated by exposure to environmental toxins such as tobacco smoke. A combination of induced enzymes may place a person at increased risk of developing lung cancer by converting precursors to carcinogens. Normal human breath contains a large number of VOCs that are endogenous and exogenous in origin, and an altered pattern of cytochrome p450 mixed oxidase activity could potentially modulate catabolism of these VOCs, thereby generating an abnormal pattern of breath VOCs.

an early warning of potentially curable disease. However, evidence from randomized trials is not yet available concerning morbidity and mortality following early detection of lung cancer. Screening for lung cancer with the breath test employing the WDA algorithm reported in this study could provide an early warning test that is safe, accurate, non-invasive, rapid, simple, and inexpensive.

WDA appears to provide a useful new technique for non-linear multivariate analysis of data. In this study, the algorithm identified dependencies beyond traditional linear approaches. WDA is a completely digital approach in the sense that it employs hard cutoff values. In future studies, multivariate modeling of clinical data could potentially produce superior results by employing a combination of linear and digital methods. A natural extension of WDA would be to investigate combinations of linear and non-linear discriminatory functions and evaluate their applicability. We conclude that WDA of breath VOCs provided a rational and accurate predictor of primary lung cancer. Because this test identifies persons at high proba-

bility for having lung cancer, it is tempting to speculate that those individuals would especially benefit from subsequent chest CT screening.

#### 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] Ralph A, Jacups S, McGough K, McDonald M, Currie BJ. The challenge of acute rheumatic fever diagnosis in a high-incidence population: a prospective study and proposed guidelines for diagnosis in Australia's Northern Territory. *Heart Lung Circ* 2006;15:113–8.
- [2] Schneider JPG, Bitterlich N, Neu K, Velcovsky HG, Morr H, Katz NEE. 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* 2003;23:899–906.
- [3] Phillips M, Gleeson K, Hughes JM, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet* 1999;353:1930–3.
- [4] Phillips M, Cataneo RN, Cummin AR, et al. Detection of lung cancer with volatile markers in the breath. *Chest* 2003;123:2115–23.
- [5] Phillips M, Cataneo R, Ditkoff B, et al. Prediction of breast cancer using volatile biomarkers in the breath. *Breast Cancer Res Treat* 2006 (e-print prior to publication).
- [6] Phillips M, Cataneo R, Condos R, et al. Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberc (Edinb)* 2007;87:44–52.
- [7] Phillips M, Boehmer J, Cataneo R, et al. Heart allograft rejection: detection with breath alkanes in low levels (the HARDBALL study). *J Heart Lung Transplant* 2004;23:701–8.
- [8] Phillips M, Altorki N, Austin JH, et al. Prediction of lung cancer using volatile biomarkers in breath. *Cancer Biomarkers* 2007;3:95–109.
- [9] Phillips M, Herrera J, Krishnan S, Zain M, Greenberg J, Cataneo RN. Variation in volatile organic compounds in the breath of normal humans. *J Chromatogr B Biomed Sci Appl* 1999;729:75–88.
- [10] Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;365:488–92.

**Table 2**  
Comparative advantages of WDA

*Intuitive* – WDA is modeled on the reasoning process that a physician employs in diagnosing a disease: it digitizes every variable and gives it a weight. This procedure is substantially simpler and more intuitive than multivariate data analysis with fuzzy logic, pattern recognition, or neural networks.

*Explicit* – WDA employs an unambiguous step-by-step procedure that generates results with clearly defined cutoff values. In contrast, multivariate data analysis with fuzzy logic, pattern recognition, or neural networks usually requires proprietary software whose inner workings are not disclosed by the provider, and may generate results that are difficult to understand.

*Rapid* – WDA is a very fast algorithm that can be readily implemented in spreadsheet programs such as Microsoft Excel.

*Superior to multilinear regression analysis in non-linear models* – WDA yielded superior results to multilinear regression, suggesting a strong non-linear component in the underlying pathophysiologic mechanism of the breath test for lung cancer.

*Direct correlation with independent variable* – WDA is mathematically monotonous in the sense that a higher gradient correlates with a higher likelihood of a particular condition e.g., cancer. This is unlike fuzzy logic, in which, for example, gradient measurements below 2 may indicate no-cancer, from 2 to 4 indicates cancer and from 4 to 8 indicates no-cancer again. The underlying pathophysiologic mechanism of the breath test for lung cancer appears to be monotonous, though non-linear.

- [11] Sone S, Li F, Yang Z, et al. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br J Cancer* 2001;84:25–32.
- [13] Hoos A, Hepp H, Kaul S, Ahlert T, Bastert G, Wallwiener D. Telomerase activity correlates with tumor aggressiveness and reflects therapy effect in breast cancer. *Int J Cancer (Pred Oncol)* 1998;79:8–12.
- [14] Sharkey JCS, Behar RJ, Perez R, et al. Minimally invasive treatment for localized adenocarcinoma of the prostate: review of 1048 patients treated with ultrasound-guided palladium-103 brachytherapy. *J Endourol* 2000;14:343–50.
- [15] Ornstein D, Rayford W, Fusaro V, 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* 2004;172:1302–5.
- [16] Bergen Hr, Vasmatis G, Cliby W, Johnson K, Oberg A, Muddiman D. Discovery of ovarian cancer biomarkers in serum using NanoLC electrospray ionization TOF and FT-ICR mass spectrometry. *Dis Markers* 2003–2004;19:239–49.
- [17] Kalgutkar AS, Hatch HL, Kosea F, et al. Preclinical pharmacokinetics and metabolism of 6-(4-(2,5-difluorophenyl)oxazol-5-yl)-3-isopropyl-[1,2,4]-triazolo [4,3-a]p yridine, a novel and selective p38alpha inhibitor: identification of an active metabolite in preclinical species and human liver microsomes. *Biopharm Drug Dispos* 2006;27:371–86.
- [18] Mortensen B, Zahlsen K, Nilsen OG. Metabolic interaction of n-hexane and methyl ethyl ketone in vitro in a head space rat liver S9 vial equilibration system. *Pharmacol Toxicol* 1998;82:67–73.
- [19] Hamuro Y, Molnar KS, Coales SJ, Ouyang B, Simorellis AK, Pochapsky TC. Hydrogen-deuterium exchange mass spectrometry for investigation of backbone dynamics of oxidized and reduced cytochrome P450(cam). *J Inorg Biochem* 2008;102:364–70.
- [20] Takemoto K, Yamazaki H, Nakajima M, Yokoi T. Genotoxic activation of benzophenone and its two metabolites by human cytochrome P450 s in SOS/umu assay. *Mutat Res* 2002;519:199–204.
- [21] Vennerstrom JL, Ager Jr AL, Andersen SL, et al. Assessment of the antimalarial potential of tetraoxane WR 148999. *Am J Trop Med Hyg* 2000;62:573–8.
- [22] Puatanachokchai R, Morimura K, Wanibuchi H, et al. Alpha-benzene hexachloride exerts hormesis in preneoplastic lesion formation of rat hepatocarcinogenesis with the possible role for hepatic detoxifying enzymes. *Cancer Lett* 2006;240:102–13.
- [23] Patrauchan MA, Florizone C, Dosanjh M, Mohn WW, Davies J, Eltis LD. Catabolism of benzoate and phthalate in *Rhodococcus* sp. strain RHA1: redundancies and convergence. *J Bacteriol* 2005;187:4050–63.
- [24] Bertea CM, Schalk M, Karp F, Maffei M, Croteau R. Demonstration that menthofuran synthase of mint (*Mentha*) is a cytochrome P450 monooxygenase: cloning, functional expression, and characterization of the responsible gene. *Arch Biochem Biophys* 2001;390:279–86.
- [25] Urlacher VB, Makhsumkhanov A, Schmid RD. Biotransformation of beta-ionone by engineered cytochrome P450 BM-3. *Appl Microbiol Biotechnol* 2006;70:53–9.
- [26] A. Jemal, R.C. Tiwari, T. Murray, et al. *Cancer Stat*, 2004. *CA Cancer J Clin* 2004;54:8–29.
- [27] Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006;355:1763–71.
- [28] Mascalchi M, Belli G, Zappa M, et al. Risk-benefit analysis of X-ray exposure associated with lung cancer screening in the Italung-CT trial. *AJR* 2006;187:421–9.