Lung Cancer 90 (2015) 92–97

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Breath carbonyl compounds as biomarkers of lung cancer

Mingxiao Li[a], Dake Yang[b], Guy Brock[b], Ralph J. Knipp[c], Michael Bousamra[d,e], Michael H. Nantz[c], Xiao-An Fu[a,*]

[a] Department of Chemical Engineering, University of Louisville, Louisville, KY 40292, United States
[b] Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY 40292, United States
[c] Department of Chemistry, University of Louisville,Louisville, KY 40292, United States
[d] Department of Cardiothoracic Surgery, University of Louisville, Louisville, KY 40292, United States
[e] James Graham Brown Cancer Center, University of Louisville, Louisville, KY 40292, United States

ARTICLE INFO

Article history:
Received 13 February 2015
Received in revised form 15 June 2015
Accepted 12 July 2015

Keywords:
Lung cancer
Exhaled breath
Biomarker
Carbonyl compound
Statistical model

ABSTRACT

Objective: Lung cancer dysregulations impart oxidative stress which results in important metabolic products in the form of volatile organic compounds (VOCs) in exhaled breath. The objective of this work is to use statistical classification models to determine specific carbonyl VOCs in exhaled breath as biomarkers for detection of lung cancer.

Materials and methods: Exhaled breath samples from 85 patients with untreated lung cancer, 34 patients with benign pulmonary nodules and 85 healthy controls were collected. Carbonyl compounds in exhaled breath were captured by silicon microreactors and analyzed by Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS). The concentrations of carbonyl compounds were analyzed using a variety of statistical classification models to determine which compounds best differentiated between the patient sub-populations. Predictive accuracy of each of the models was assessed on a separate test data set.

Results: Six carbonyl compounds (C_4H_8O, $C_5H_{10}O$, $C_2H_4O_2$, $C_4H_8O_2$, $C_6H_{10}O_2$, $C_9H_{16}O_2$) had significantly elevated concentrations in lung cancer patients vs. controls. A model based on counting the number of elevated compounds out of these six achieved an overall classification accuracy on the test data of 97% (95% CI 92%–100%), 95% (95% CI 88%–100%), and 89% (95% CI 79%–99%) for classifying lung cancer patients vs. non-smokers, current smokers, and patients with benign nodules, respectively. These results were comparable to benchmarking based on established statistical and machine-learning methods. The sensitivity in each case was 96% or higher, with specificity ranging from 64% for benign nodule patients to 86% for smokers and 100% for non-smokers.

Conclusion: A model based on elevated levels of the six carbonyl VOCs effectively discriminates lung cancer patients from healthy controls as well as patients with benign pulmonary nodules.

1. Introduction

Lung cancer is the leading cause of cancer mortality. Each year more people die of lung cancer than of breast, colon, and prostate cancers combined [1]. The 5-year survival rate of a stage I patient can be as high as 60%–80%; however, only 15% of lung cancer patients are diagnosed before they develop regional or metastatic disease [2,3]. The National Lung Screening Trial Team recently reported that computed tomography (CT) screening of heavy smokers reduced lung cancer deaths by 20% [4]. Therefore, detection of lung cancer in its early stages is critical for increasing patient survival [5].

One promising method for detection of early lung cancer involves analysis of volatile organic compounds (VOCs) in exhaled breath [5–12]. Growth of cancer cells in an unstable redox environment is thought to cause oxidative stress [13] and the production of detectable VOCs in exhaled breath. Solid phase micro-extraction–gas chromatography mass spectrometry (SPME–GCMS) has been used to identify lung cancer biomarkers in exhaled breath [5,14,15]. However, SPME is based on physical adsorption and the approach can only provide a semi-quantitative analysis of breath VOCs. Exhaled breath condensate (EBC) has also been a popular method for breath analysis [16]. One limitation of EBC is the variable dilution of moisture in exhaled breath that

* Corresponding author.
E-mail address: xiaoan.fu@louisville.edu (X.-A. Fu).

http://dx.doi.org/10.1016/j.lungcan.2015.07.005
0169-5002/

may contribute to large variability in concentration measurements. Finally, while electronic noses (eNose) have the advantage of fast detection of VOCs [7,17,18], the inability to identify specific VOCs and the interference of matrix VOCs in exhaled breath limits their application.

Volatile ketones and aldehydes can be generated in vivo by lipid peroxidation [19,20]. We have developed a technique utilizing quaternary aminooxy-coated silicon microreactors for selective capture and quantification of these ketones and aldehydes in air [21–23] and exhaled breath [24,25]. In this present work, a variety of statistical classification models were used to analyze how the measured concentrations of carbonyl compounds can be used to determine which compounds and models best differentiate between the patient sub-populations.

2. Materials and methods

2.1. Materials

All reagents and solvents, including deuterated acetone (acetone-d6) (99.9%), and methanol (99.9%), were purchased from Sigma–Aldrich. The quaternary ammonium aminooxy compound 2-(aminooxy) ethyl-*N*,*N*,*N*-trimethylammonium iodide (ATM) was synthesized according to a published method [26]. One liter Tedlar bags were also purchased from Sigma–Aldrich. Tedlars bags and syringes were tested free of carbonyl compound contamination.

2.2. Test population, breath sampling and analysis

Exhaled breath samples of 85 patients with untreated lung cancer, 34 patients with benign pulmonary nodules and 85 healthy control patients (40 nonsmokers and 45 current smokers) were collected and analyzed. The detailed research protocol for collection of exhaled breath samples was approved by the Institutional Review Board (IRB). The healthy controls were recruited from members of the general population who are free of lung cancer or any other chronic pulmonary disease. All patients had pulmonary nodules evidenced by computed tomography (CT) scans and were recruited at the James Graham Brown Cancer Center at the University of Louisville. The diagnostic conclusions from these breath analyses were confirmed by histopathology, or in some cases of benign disease, by radiographic stability or resolution while followed over a two year period. Table 1 lists the study subject information.

Each volunteer filled a 1 L Tedlar bag with breath through one exhalation to provide a mixture of alveolar and tidal breath. Ambient air samples (1 L) were also collected in the clinic room to serve as a reference of carbonyl compounds. After collection of breath or air samples, Tedlar bags were connected to the silicon microreactors through deactivated silica tubes as shown in the schematic setup of Figure S1 (supplementary). The breath samples flow from the Tedlar bags through the microreactor at a flow rate of 5 mL/min by applied vacuum. The surfaces of the micropillars in the microreactors are functionalized by ATM, for capture of aldehydes and ketones via oximation reactions [21–23]. After complete deflation of the sample bags, ATM and its adducts were eluted from the microreactor by flowing methanol (100 μL) through the microreactor [21]. A known amount of ATM–acetone-d6 adduct in methanol was added to each sample to serve as an internal reference of FT-ICR-MS to determine the amounts of detected carbonyl VOCs [21,22].

The eluent solutions were analyzed using a hybrid linear ion trap-FT-ICR-MS instrument (Finnigan LTQ-FT, Thermo Electron, Bremen, Germany) equipped with a TriVersa NanoMate ion source (Advion BioSciences, Ithaca, NY) with a nanoelectrospray chip (nozzle inner diameter 5.5 μm).

2.3. Statistical data analysis

Since much research has indicated that tobacco smoking can affect the VOCs in exhaled breath [5,14], the measured carbonyl VOCs from this study were grouped according to the following categories of patients: lung cancer patients, patients with benign pulmonary nodules, controls who have never smoked, and controls who are current smokers. Differences in VOC concentrations between the four groups were visualized using boxplots and assessed for statistically significant differences using the non-parametric Kruskal–Wallis test to count for the non-normal nature of the distributions.

Next, we endeavored to build several classification models comparing (a) lung cancer patients with controls who have never smoked, (b) lung cancer patients with current smoking controls, and (c) lung cancer patients with patients having benign pulmonary nodules. The data for each comparison were first split randomly into a training (70% of the data) and test (30% of the data) set. Several models were fitted to the training data including (a) partial least squares (PLS), (b) support vector machines (SVM), (c) random forest (RF), (d) linear discriminant analysis (LDA), and (e) quadratic discriminant analysis (QDA). To determine the optimal set of VOCs for classification for each model, a wrapper method called recursive feature elimination (RFE) was used for variable selection. RFE is a procedure that ranks variables based on a variable importance (VI) score. The VI score is the area under the receiver operating characteristic (ROC) curve for all methods except PLS and RF, which have their own built-in VI scores [27]. The least important variables are then removed and the classification model is re-fitted. The procedure is repeated recursively until the accuracy of the model begins to diminish. The leave group out cross-validation (LGOCV) approach was used for RFE and also to optimize the tuning parameters for each method (e.g., the number of PLS components). After this, the models fitted above were used to predict the patient classes in the test data set. The sensitivity, specificity, and accuracy of each model was calculated and compared. The R packages caret [27] and pROC [28] were used for all calculations.

The aforementioned machine learning methods have repeatedly demonstrated good classification performance in the literature and hence are useful for benchmarking purposes. However, they can be difficult to interpret and implement in clinical practice. Therefore, we sought to fit a simpler classification model with similar performance to the methods listed above. A combined set of VOCs was determined from the best performing classification models, and an optimal classification cut-point for each VOC was determined from the ROC curve fitted to the training data based on Youden's index [29]. Since lung cancer patients had higher concentrations of each VOC compared to the other subjects, each VOC was scored as zero or one depending on whether the level was below or above the threshold, respectively. The VOC scores were then added for each subject to obtain a total VOC score, and a logistic regression model was fitted for each of the three comparisons (lung cancer vs. benign nodule, lung cancer vs. current smokers, and lung cancer vs. non-smokers) using the total VOC score as the predictor variable. In addition, we evaluated the performance of this simplified VOC model for lung cancer vs. all other subjects and lung cancer vs. benign nodules and an age-restricted (age $\geq$50) set of control patients. The latter was used to account for the potential age discrepancy between lung cancer and control patients. A classification cut-off based on the total VOC score was determined based on the predicted probability threshold of 0.5. Sensitivity, specificity, and overall accuracy were then evaluated in the test set based on this total VOC score threshold, with 95% confidence intervals calculated using either Wald's method or, for values equal to 1, the score method [30].

Table 1
Study subject information.

Demographics		Lung cancers ($N = 85$)	Benign nodules ($N = 34$)	Healthy controls ($N = 85$)
Male (N, %)		46 (54%)	10 (29%)	43 (50%)
Age (mean ± SD)		66.12 ± 10.1	51.71 ± 15.3	42.15 ± 14.2
Smoking history (N, %)	Current	45 (53%)	10 (29%)	45 (53%)
	Former	34 (40%)	7 (21%)	
	Never	2 (2%)	7 (21%)	40 (47%)
	Missing	4 (5%)	10 (29%)	
Etiology				
Stage (N, %)	I	30 (35%)		
	II	14 (16%)		
	III	21 (25%)		
	IV	14 (16%)		
	Unknown	6 (7%)		
Type (N, %)	NSCLC	78 (92%)		
	Adenocarcinoma	31 (36%)		
	Squamous cell	33 (39%)		
	Other	14 (16%)		
	SCLC	7 (8%)		
Benign Nodules (N, %)	Pneumonia		2 (6%)	
	Granuloma		8 (24%)	
	Inflammation		3 (9%)	
	Histoplasmosis		2 (6%)	
	Epithelial cells		2 (6%)	
	Foamy macrophages		1 (3%)	
	Unknown (radiographically benign)		16 (47%)	

3. Results

Carbonyl VOCs from one carbon atom (formaldehyde) to ten carbon atoms were routinely detected in exhaled breath samples of healthy controls and patients. As determined by FT-ICR-MS, molecular formulas of the VOCs include CH_2O, C_2H_4O, $C_2H_4O_2$, C_3H_4O, $C_3H_4O_2$, C_3H_6O, C_4H_8O, $C_4H_8O_2$, $C_5H_{10}O$, $C_6H_{12}O$, $C_6H_{10}O_2$, $C_7H_{14}O$, $C_8H_{16}O$, $C_9H_{18}O$, $C_9H_{16}O_2$, and $C_{10}H_{20}O$. The FT-ICR-MS spectra in Figure S2 (Supplementary) show typical relative abundances among these VOCs in breath samples of lung cancer patients and healthy controls. Constitutionally isomeric ketones and aldehydes are indistinguishable by direct infusion one dimensional FTICR-MS, however, the measured molecular weight at a resolving power of 200,000 provides accurate chemical formulas. Separation and structure identification of some important isometric ketones and aldehydes was done by FT-ICR-MS/MS and GC–MS.

Statistically significant differences between the four patient groups (lung cancer patients, patients with benign nodules, current smoking controls, and non-smoking controls) were found for nine carbonyl VOCs ($p < 10^{-4}$ based on Kruskal–Wallis test). Fig. 1 shows boxplots of the concentrations of these nine VOCs of the four groups. These compounds are C_4H_8O: 2-butanone; $C_4H_8O_2$: 3-hydroxy-2-butanone; $C_6H_{10}O_2$: 4-hydroxy-2-hexenal (4-HHE); $C_2H_4O_2$: hydroxyacetaldehyde; C_3H_4O: acrolein; $C_9H_{16}O_2$: 4-hydroxy-2-nonenal (4-HNE); $C_2H_4O_2$: acetaldehyde; $C_3H_4O_2$: malondialdehyde (MDA); and $C_5H_{10}O$: a mixture of 2-pentanone and pentanal. Other carbonyl VOCs were not statistically different among the four patient groups. Several carbonyl VOCs (e.g., formaldehyde: CHO, acetaldehyde: C_2H_4O, acetone: C_3H_6O) were elevated in smoking controls and patients with benign nodules, indicating these VOCs would not be good for differentiating these two groups from lung cancer patients. Figs. 2, 3, and 4 show ROC curves based on the entire patient population for the three comparisons of interest (lung cancer vs. benign nodule, lung cancer vs. current smokers, and lung cancer vs. non-smokers, respectively) indicated that the three strongest single classifiers in each case were $C_4H_8O_2$, C_4H_8O, and $C_5H_{10}O$. While a single marker appears to achieve excellent discrimination between lung cancer patients and controls (highest AUC of 0.962 and 0.946 vs. non-smokers and smokers, respectively), this is less so for lung cancer patients compared to patients with benign nodules (highest AUC of 0.901).

Table 2
Classification threshold concentrations (nmol/L) for selected VOCs. Classification threshold concentations for each compound to classify lung cancer (LC) patients vs. patients with benign nodules (BN), current smoking controls (Sm), and non-smoking controls (NS). Threshold concentrations were based on Youden's index using the training data.

VOC	LC vs. BN	LC vs. Sm	LC vs. NS
C_4H_8O	2.36	2.365	2.255
4-HHE	0.0073	0.00734	0.00672
$C_4H_8O_2$	0.1695	0.1695	0.1595
$C_2H_4O_2$	0.357	0.3575	0.3875
4-HNE	0.00175	0.000285	0.000255
$C_5H_{10}O$	1.1	1.107	1.315

Classification results from the five fitted models (PLS, SVM, RF, QDA, and LDA) for each of the three comparisons on the test data is given in Table S1 (supplementary), with the sensitivity and specificity of the best performing models for each comparison given in Table S2. The final set of selected carbonyl markers for each of the top-performing models is given in Table S3. The selected VOCs generally agree with the best single compounds, though there are some differences (e.g., $C_2H_4O_2$ and 4-HHE) selected for the PLS model for lung cancer patients vs. patients with benign nodules, both of which have individual AUC values outside the top three for this comparison).

The top-performing models in Table S2 provide a benchmark for classification performance based on the VOCs. To build a simpler, easy to implement classifier we selected a set of VOCs from Table S3 by taking the union of all the listed VOCs (C_4H_8O, 4-HHE, $C_4H_8O_2$, $C_2H_4O_2$, 4-HNE, and $C_5H_{10}O$). Optimal classification threshold concentrations for each of these VOCs based on Youden's index (calculated on the training data only) are given in Table 2. As described in the methods, a total score for each patient was obtained by summing the number of VOCs above these thresholds (possible values of 0–6). These total VOC scores were then used in a logistic regression analysis to determine thresholds for classifying lung cancer vs. controls and patients. For all three comparisons a total VOC score of two or more was classified as cancer. The sensitivity, specificity, and overall accuracy for lung cancer vs. benign

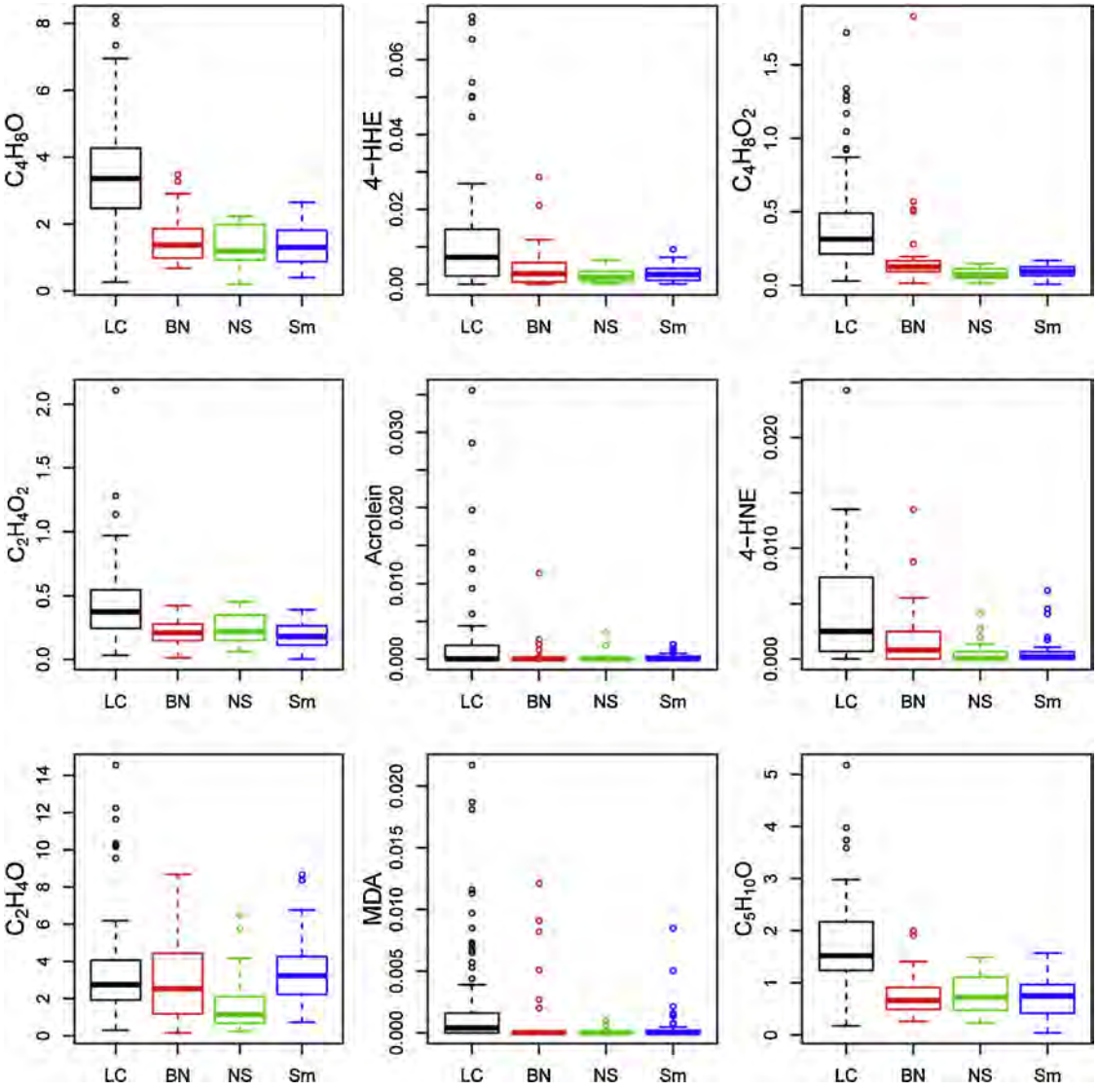

Fig. 1. Boxplots of the concentrations (nmol/L) of nine VOCs evaluated in this study, stratified by lung cancer patients (LC), patients with benign nodules (BN), non-smoking controls (NS), and current smoking controls (Sm).

nodules on the test data were 26/26 (1.00, 95% CI 0.94–1.00), 7/11 (0.64, 95% CI 0.35–0.92), and 33/37 (0.89, 95% CI 0.79–0.99) respectively, while for lung cancer vs. current smoking controls it was 26/26 (1.00, 95% CI 0.94–1.00), 12/14 (0.86, 95% CI 0.67–1.00), and 38/40 (0.95, 95% CI 0.88–1.00) and for lung cancer vs. non-smoking controls it was 25/26 (0.96, 95% CI 0.89–1.00), 12/12 (100%, 95% CI 0.88–1.00), and 37/38 (97%, 95% CI 0.92–1.00). These values compare favorably to the optimal classification performances given in Table S2 for each comparison.

To further check the performance of the VOC classifier as a population screening test for lung cancer, we combined all of the test samples and used the thresholds determined for LC vs. smokers and LC vs. non-smokers (Table 2) to classify patients. As before, a total VOC score (number of elevated VOCs) of two or more was classified as cancer. The sensitivity, specificity, and overall accuracy in the test data for lung cancer patients vs. all other groups were 25/26 (0.96, 95% CI 0.88–1.00), 31/37 (0.84, 95% CI 0.72–0.96), and 56/63 (0.89, 95% CI 0.81–0.97), respectively. Next, we assessed performance of the classifier based on an additional set of 70 age-restricted controls combined with the benign nodule patients, using the same thresholds. These age-restricted controls subjects had a mean age of 62.9, with a median of 64 and a standard deviation of 8.1, similar to the lung cancer patients. Among these controls 13 (18.6%) were current smokers, 27 (38.6%) non-smokers, and 30 (42.9%) former smokers (for these subjects we used the LC vs. non-smoker thresholds). The sensitivity was the same, while the specificity and overall accuracy for this group of patients and controls were 64/81 (0.79, 0.70–0.88) and 89/107 (0.83, 95% CI 0.76–0.90), respectively, which was similar albeit slightly lower compared to the non-age-restricted results.

4. Discussion

This study demonstrates that the six markers in exhaled breath can be used for distinguishing patients with lung cancer from patients with benign pulmonary nodules and healthy subjects. The simple model based on counting the number of elevated markers achieved high sensitivity and specificity — comparable to that from the more complicated statistical/machine-learning models. This study has several significant advantages related to the microfabricated microreactors and the highest resolution of FT-ICR-MS. First, the microrectors are designed with thousands of micropillars to provide high capture efficiencies of carbonyl compounds in breath [21–23]. Second, chemoselective capture of carbonyl compounds

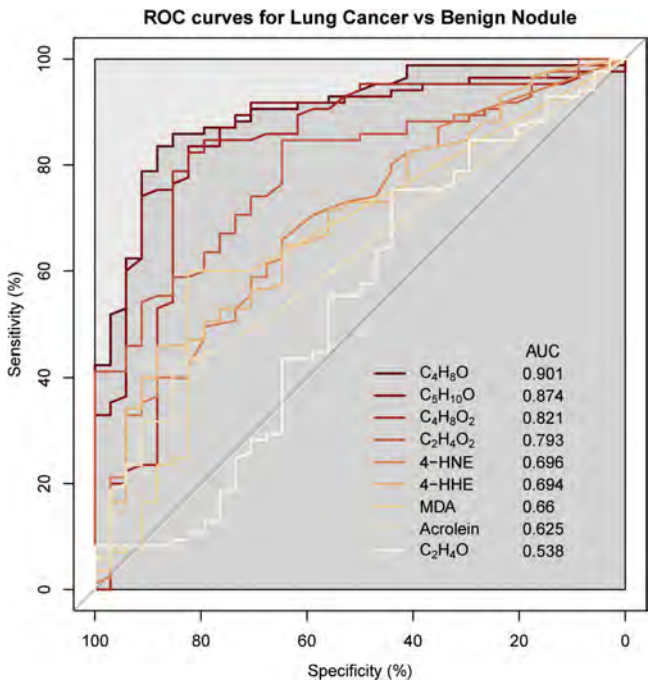

Fig. 2. ROC curves corresponding to the nine evaluated VOCs for discriminating lung cancer patients vs. patients with benign nodules. VOCs are sorted in order of decreasing area under the curve (AUC) values.

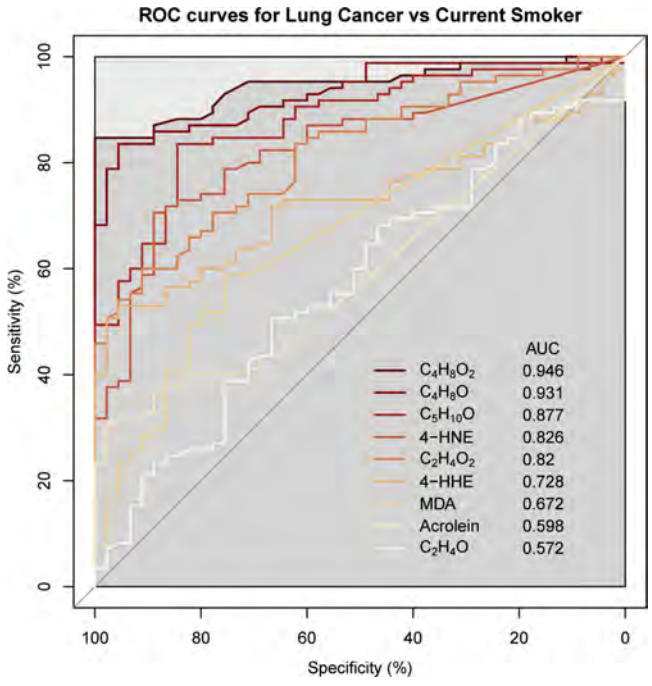

Fig. 3. ROC curves corresponding to the nine evaluated VOCs for discriminating lung cancer patients vs. controls who are current smokers. VOCs are sorted in order of decreasing area under the curve (AUC) values.

through aminooxy reactions simplifies the spectrum of compounds to be quantitated.

Diet and smoking are known to influence the relative composition of breath [31,32]. Although no diet controls were used in this study, there was no notable change for the results of analysis of carbonyl compounds related to recent oral intake. Therefore, the identification of the six carbonyl compounds is unlikely affected by diet. Ambient air also contains many volatiles that are present in

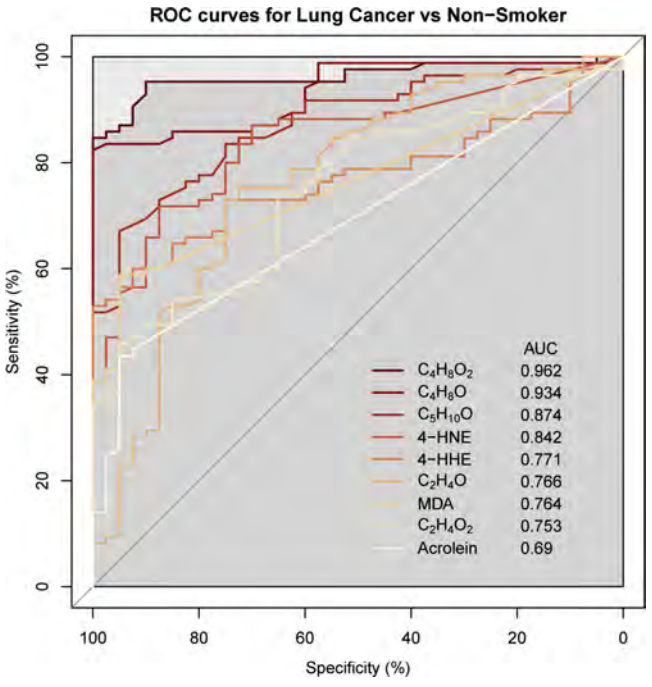

Fig. 4. ROC curves corresponding to the nine evaluated VOCs for discriminating lung cancer patients vs. non-smoking controls. VOCs are sorted in order of decreasing area under the curve (AUC) values.

the breath [33]. The concentrations of the six carbonyls in ambient air of the clinical room were much lower than exhaled breath and did not affect the diagnosis results.

Our previous work identified four lung cancer markers C_4H_8O, 4-HHE, $C_4H_8O_2$, and $C_2H_4O_2$ [24] and a diagnostic method based on counting the number of elevated these four markers was developed [25]. In this study, two additional markers 4-HNE and $C_5H_{10}O$ were identified by the statistical classification models. We compared the performance of the six and four marker models based on the test data sets in this study, using the same thresholds from Table 2 and a threshold of one elevated VOC classified as cancer. For the comparison between lung cancer cases and benign nodules, both the four and six marker models achieved 100% sensitivity but the four marker model had lower specificity (55% vs 64%). However, it should be noted that the specificity estimates are rather imprecise as they are based on only 11 test data subjects. For the lung cancer vs. smoking controls comparison the models achieved identical results, while for the cancer vs. non-smokers comparison the four marker model had slightly higher sensitivity (100% vs. 96%) and slightly lower specificity (92% vs. 100%). In the latter case both models misclassified only a single patient. Collectively, the results suggest that the simple four marker model may be sufficient for a population screening device but the six marker model might be better for discerning between lung cancer patients and patients with benign nodules. Additional advantages of the current study include benchmarking of the simpler models with established statistical/machine-learning methods and optimization of cut-points for both individual VOCs and the number of elevated VOCs for determining whether a patient has lung cancer. It should also be noted that the carbonyl VOCs were equally effective in separating lung cancer cases from both smoking and non-smoking controls, in agreement with what we previously observed [24,25]. A limitation of the current work is that we selected the six VOCs for our simple model based on the best performing classifiers in the test data. Hence information from the test data was used to determine this final marker set. Ultimately, we will validate both the six and four marker VOC models based on a much larger data set.

5. Conclusion

Six VOCs (2-butanone, 4-HHE, 3-hydroxy-2-butanone, hydroxyacetaldehyde, 4-HNE, and the $C_5H_{10}O$ combination of 2-pentanone and pentanal) were found to effectively distinguish LC from patients with benign pulmonary nodules and healthy controls. A simple model based on scoring the number of VOCs above a given threshold concentration achieves an overall classification accuracy of 89%, 95%, and 97% for classifying lung cancer patients vs. patients with benign nodules, current smoking controls, and non-smoking controls, respectively. In each case the sensitivity of the model for the three separate comparisons was at or above 96%. The specificity was lowest for lung cancer vs. patients with benign nodules (64%), but higher for non-smokers (100%), smokers (86%), and for the three groups combined (84%). The numbers for this model compare favorably with the top performing machine-learning methods we evaluated (c.f. Tables S1 and S2), indicating that this easy-to-implement model achieves near optimal performance. A study with a larger study population is required to verify the method.

Funding

This study was partially supported by the National Science Foundation under Grant CBET-1159829, the Kentucky Lung Cancer Research Program, and the University of Louisville.

Conflict of interest

X.A. Fu and M.H. Nantz hold a patent on silicon microreactors for concentrating trace carbonyl compounds in air and breath. B. Bousamra, X.A. Fu and M.H. Nantz hold a pending patent application on carbonyl compounds as biomarkers of lung cancer.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.lungcan.2015.07.005

References

[1] A.J. Alberg, J.G. Ford, J.M. Samet, American college of chest physicians epidemiology of lung cancer: AACP evidence-based clinical practice guidelines (2nd edition), Chest 132 (2007) 29S–55S.
[2] L. Dominioni, A. Imperatori, F. Rovera, A. Ochetti, G. Torrigiotti, M.I. Paolucci Stage, Nonsmall cell lung carcinoma: analysis of survival and implications for screening, Cancer 89 (2000) 2334–2344.
[3] A. Jemal, R. Siegel, J. Xu, E. Ward, Cancer statistics, CA Cancer J. Clin. 60 (2011) 277–300.
[4] D.R. Aberle, A.M. Adams, C.D. Berg, W.C. Black, J.D. Clapp, et al., Reduced lung-mortality with low-dose computed tomographic screening, New Engl. J. Med. 365 (5) (2011) 395–409.
[5] N. Savage, Early detection: spotting the first signs, Nature 471 (2011) S14–S15.
[6] A. Bajtarevic, C. Ager, M. Pienz, M. Klieber, L. Schwarz, M. Ligor, T. Ligor, W. Filipiak, H. Denz, M. Fiegl, W. Hilbe, W. Weiss, P. Lukas, H. Jamnig, M. Hackl, A. Haidenberger, B. Buszewski, W. Miekisch, J. Schubert, A. Amann, Noninvasive detection of lung cancer by analysis of exhaled breath, BMC Cancer 9 (2009) 348–363.
[7] G. Peng, U. Tisch, U. Adams, M. Hakim, N. Shehada, Y.Y. Broza, S. Billan, R. Abdah-Bortnyak, A. Kuten, H. Haick, Diagnosing lung cancer in exhaled breath using gold nanoparticles, Nature Nanotechnol. 4 (2009) 669–673.
[8] X. Chen, F. Xu, Y. Wang, Y. Pan, D. Lu, P. Wang, K. Ying, E. Chen, W.A. Zhang, Study of the volatile organic compounds exhaled by lung cancer cells In vitro for breath diagnosis, Cancer 110 (2007) 835–844.
[9] A. Wehinger, A. Schmid, S. Mechtcheriakov, M. Ledochowski, C. Grabmer, G.A. Gastl, A. Amann, Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas, InternJ. Mass Spectr. 265 (2007) 49–59.
[10] M. Phillips, K. Gleeson, J.M.B. Hughes, J. Greenberg, R.N. Cataneo, L. Baker, P. McVay, Volatile organic compounds in breath as markers of lung cancer: a crosssectional study, Lancet 353 (1999) 1930–1933.
[11] G.L.J. Preti, J.G. Kostelc, S. Aldinger, R. Daniele, Analysis of lung air from patients with bronchogenic carcinoma and controls using gas chromatography-mass spectrometry, J. Chromatogr. 432 (1988) 1–11.
[12] S.M. Gordon, J.P. Szidon, B.K. Krotoszynski, R.D. Gibbons, H.J. O'Neill, Volatile organic compounds in exhaled air from patients with lung cancer, Clin. Chem. 31 (1985) 1278–1282.
[13] M. Hakim, Y.Y. Broza, O. Barash, N. Peled, M. Phillips, A. Amann, H. Haick, Volatile organic compounds of lung cancer and possible biochemical pathways, Chem. Rev. 112 (2012) 5949–5966.
[14] Y. Wang, Y. Hu, D. Wang, K. Yu, L. Wang, Y. Zou, C. Zhao, X. Zhang, P. Wang, K. Ying, The analysis of volatile organic compounds biomarkers for lung cancer in exhaled breath, tissues and cell lines, Cancer Biomark. 11 (2012) 129–137.
[15] D. Poli, P. Carbognani, M. Corradi, M. Goldoni, O. Acampa, B. Balbi, L. Bianchi, M. Rusca, A. Mutti, Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study, Respir. Res. 6 (2005) 71–81.
[16] H.P. Chan, V. Tran, C. Lewis, P. Thomas, Markers of oxidative stress in exhaled breath of subjects with lung cancer, Respirology 13 (2008) A58.
[17] P.J. Mazzone, X. Wang, Y. Xu, T. Mekhail, M.C. Beukemann, J. Na, J.W. Kemling, K.S. Suslick, M. Sasidhar, Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer, J. Thorac. Oncol. 7 (2012) 137–142.
[18] S. Dragonieri, J.T. Annema, R. Schot, M.P.C. van der Schee, A. Spanevello, P. Carratu, O. Resta, K.F. Rabe, E.H. Bel, P.J. Sterk, An electronic nose in the discrimination of paitents with non-small cell ung cancer and COPD, Lung Cancer 64 (2009) 166–170.
[19] E.K. Long, M.J. Picklo, Trans-4-hydroxy-2-hexenal, a product of n-3 fatty acid peroxidation: make some room HNE. . ., Free Radic. Biol. Med. 49 (2010) 1–8.
[20] Y. Riahi, G. Cohen, O. Shamni, S. Sasson, Signaling and cytotoxic functions of 4-hydroxyalkenals, Am. J. Physiol. Endocrinol. Metab. 299 (2010) E879–E886.
[21] M. Li, S. Biswas, M.H. Nantz, R.M. Higashi, X. Fu, Preconcentration and analysis of trace volatile carbonyl compounds, Anal. Chem. 84 (2012) 1288–1293.
[22] X. Fu, M. Li, S. Biswas, M.H. Nantz, R.M. Higashi, A novel microreactor approach for analysis of ketones and aldehydes in breath, Analyst 136 (2011) 4662–4666.
[23] M. Li, S. Biswas, M.H. Nantz, R.M. Higashi, X.A. Fu, A microfabricated preconcentration device for breath analysis, Sens. Act. B 180 (2013) 130–136.
[24] M. Bousamra, E. Schumer, M. Li, R.J. Knipp, M.H. Nantz, V.V. Berkel, X.A. Fu, Quantitative analysis of exhaled carbonyl compounds distinguishes benign from malignant pulmonary disease, J. Thoracic. Cardio. Surg. 148 (2014) 1074–1081.
[25] X.A. Fu, M. Li, R.J. Knipp, M.H. Nantz, M. Bousamra, Noninvasive detection of lung cancer using exhaled breath, Cancer Med. 3 (2014) 174–181.
[26] S. Biswas, X. Huang, W.R. Badger, M.H. Nantz, Nucleophilic cationization reagents, Tetrahedron Lett. 51 (2010) 1727–1729.
[27] M. Kuhn, Building predictive models in R using the caret package, J. Stat. Software 28 (2008) 5.
[28] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, M. Müller, pROC: an open-source package for R and S+ to analyze and compare ROC curves, BMC Bioinf. 12 (2011) 77.
[29] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (1950) 32–35.
[30] A. Agresti, Categorical Data Analysis, 3rd ed., Wiley, Hoboken, New Jersey, 2013, pp. 4–17.
[31] W. Miekisch, S. Kischkel, A. Sawacki, T. Liebau, M. Mieth, J.K. Schubert, Impact of sampling procedures on the results of breath analysis, J. Breath Res. 2 (2008) 26007.
[32] K.A. Cope, M.T. Watson, W.M. Foster, S.S. Sehnert, T.H. Risby, Effects of ventilation on the collection of exhaled breath in humans, J. Appl. Physiol. 97 (2004) 1371–1379.
[33] S. Dragonieri, R. Schot, B.J.A. Mertens, S. Le Cessie, S.A. Gauw, A. Spanevello, O. Resta, N.P. Willard, T.J. Vink, K.F. Rabe, E.H. Bel, P.J. Sterk, An electronic nose in the discrimination of patients with asthma and controls, J. Allergy Clin. Immunol. 120 (2007) 856–862.