Progress in the Development of Volatile Exhaled Breath Signatures of Lung Cancer

Peter J. Mazzone¹, Xiao-Feng Wang¹, Sung Lim², James Jett³, Humberto Choi¹, Qi Zhang¹, Mary Beukemann¹, Meredith Seeley¹, Ray Martino², and Paul Rhodes²

¹Respiratory Institute, Cleveland Clinic, Cleveland, Ohio; ²Metabolomx, Mountain View, California; and ³National Jewish Health, Denver, Colorado

Abstract

Rationale: Volatile organic compounds present in the exhaled breath have shown promise as biomarkers of lung cancer. Advances in colorimetric sensor array technology, breath collection methods, and clinical phenotyping may lead to the development of a more accurate breath biomarker.

Objectives: Perform a discovery-level assessment of the accuracy of a colorimetric sensor array–based volatile breath biomarker.

Methods: Subjects with biopsy-confirmed untreated lung cancer, and others at risk for developing lung cancer, performed tidal breathing into a breath collection instrument designed to expose a colorimetric sensor array to the alveolar portion of the breath. Random forest models were built from the sensor output of 70% of the study subjects and were tested against the remaining 30%. Models were developed to separate cancer and subgroups from control, and to characterize the cancer. Additional models were developed after matching the clinical phenotypes of cancer and control subjects.

Measurements and Main Results: Ninety-seven subjects with lung cancer and 182 control subjects participated. The accuracies, reported as C-statistics, for models of cancer and subgroups versus control ranged from 0.794 to 0.861. The accuracy was improved by developing models for cancer and control groups selected through propensity matching for clinical variables. A model built using only subjects from the largest available clinical subgroup (49 subjects) had a C-statistic of 0.982. Models developed and tested to characterize cancer histology, and to compare early- with late-stage cancer, had C-statistics of 0.881–0.960.

Conclusions: The colorimetric sensor array signature of exhaled breath volatile organic compounds was capable of distinguishing patients with lung cancer from clinically relevant control subjects in a discovery level trial. The incorporation of clinical phenotypes into the further development of this biomarker may optimize its accuracy.

Keywords: breath testing; volatile organic compounds; phenotypes

(Received in original form November 23, 2014; accepted in final form February 12, 2015)

Supported by Metabolomx (SBIR contract HHSN261201100066C awarded to Metabolomx, and provided to the Cleveland Clinic and National Jewish Health).

Author Contributions: P.J.M. takes responsibility for the content of the manuscript, including the data and analysis. P.J.M. made substantial contributions to the conception and design of the research, acquisition of data, analysis and interpretation of data, drafting and revision of the article, and provided final approval of the version to be published. X.-F.W. and Q.Z. made substantial contributions to the analysis and interpretation of data and revision of the submitted article, have provided final approval of the version to be published and have agreed to be accountable for all aspects of the work. S.L., R.M., and P.R. made substantial contributions to the conception and design of the technical portions of the study and the revision of the manuscript in these areas, have provided approval of the version submitted and agree to be accountable for all aspects of the work. H.C., J.J., M.B., and M.S. made substantial contributions to the acquisition of data and revision of the submitted article, provided final approval of the version submitted, and agree to be accountable for all aspects of the work.

Correspondence and requests for reprints should be addressed to Peter J. Mazzone, M.D., M.P.H., 9500 Euclid Avenue, A90, Cleveland, OH 44195. E-mail: mazzonp@ccf.org

Ann Am Thorac Soc Vol 12, No 5, pp 752–757, May 2015

Copyright © 2015 by the American Thoracic Society

DOI: 10.1513/AnnalsATS.201411-540OC

Internet address: www.atsjournals.org

Accurate, noninvasive biomarkers have the potential to improve our ability to prevent, diagnose, and treat lung cancer. For these reasons there has been tremendous interest in the development of molecular biomarkers for lung cancer. Biomarkers of single or

multiple alterations in the genome, transcriptome, proteome, and metabolome are being developed. To date, only the analysis of epidermal growth factor receptor mutations and EML4–ALK (echinoderm microtubule-associated protein like-4–

anaplastic lymphoma kinase) translocations in tumor tissue have been shown to affect clinical management to the benefit of our patients (1, 2). Other technically validated biomarkers that are available or soon to be available for use include serum measures

of antibodies to tumor associated antigens (3), protein (4) and microRNA profiles (5), as well as an airway gene signature (6).

Our exhaled breath may be the least invasive biospecimen available for analysis. In addition to nitrogen, oxygen, and carbon dioxide (CO₂), our exhaled breath contains volatile organic compounds (VOCs) in low concentrations. These VOCs are believed to reflect metabolic alterations at the tissue level, the products of which enter the bloodstream and leave the body in the breath, urine, and through the skin and mucosal surfaces. Many lines of evidence support the presence of metabolic alterations in tumor tissue and the potential usefulness of breath VOC analysis. There have been reports detailing differences in the way tumor tissue handles energy stresses (7), handles oxidative stresses (8), and metabolizes specific classes of VOCs (9) when compared with healthy tissues. The composition of nonvolatile small-molecule metabolites has been found to differ in patients with lung cancer (10). VOC analysis of the headspace gas of lung cancer cell lines (11), and exhaled breath VOC profiles assessed by gas chromatography-mass spectrometry (12) and with cross-responsive chemical sensors (13), have all supported this line of research.

We have previously reported the results of exhaled breath VOC biomarker development, using a colorimetric sensor array (CSA); a CSA is a cross-responsive chemical sensor whose output is a change in the color of its chemoresponsive elements (14, 15). In the current study we report on the impact of advances in the CSA, breath collection methodology, and clinical phenotyping on the accuracy of a breath VOC biomarker of lung cancer. Some of the results of this study have been previously reported in the form of an abstract (16).

Methods

Study Subjects

This study was conducted in accordance with the amended Declaration of Helsinki. The institutional review boards of the Cleveland Clinic (CC, Cleveland, OH) (IRB 7084) and National Jewish Health Center (NJH, Denver, CO) (HS 2595) approved the protocol. All study subjects provided written informed consent.

Study subjects were included as cases if they had biopsy-confirmed, untreated lung cancer or an imaging abnormality highly suspicious for lung cancer, later confirmed to be lung cancer. Study subjects were included as control subjects if they were at risk for developing lung cancer, based on age greater than 40 years and either tobacco use of at least 10 pack-years, a family history of lung cancer, or the presence of chronic obstructive pulmonary disease (COPD) by history; or if they presented with an indeterminate lung nodule 8–30 mm in diameter. Those with lung nodules were confirmed to be free of lung cancer on the basis of biopsy, or stability on serial imaging for a timeframe in keeping with current guidelines (17). Study subjects were excluded from participation if they had a prior history of lung cancer, a history of another cancer within 5 years, were receiving immunosuppression, or were using continuous supplemental oxygen. Data collection included demographic variables and comorbidities for all subjects, nodule size for control subjects with lung nodules, and cancer histology and stage for the subjects with cancer.

CSA and Breath Collection

The CSA was designed to have 4 channels, each containing 32 sensor elements. The sensor elements were chemoresponsive pigments contained in organically modified siloxanes chosen for their combined reactivity to diverse classes of VOCs. One of the four channels contained an oxidizer to derivatize VOCs into more reactive species, and a second channel contained a desiccant in addition to the oxidizer to minimize the impact of breath humidity on the responses

of that channel (Figure 1). The sensor was housed in a nest maintained at 30°C to optimize response and minimize condensation of water vapor.

Study subjects performed tidal breathing while inhaling through a VOC filter. A 75% fall in the upward slope of the end-tidal CO₂ curve triggered the opening of a valve, leading to exposure of the CSA to the alveolar portion of the breath. Alveolar breath was drawn across the CSA by mechanical pumps positioned distal to the sensor until a total of 1 L of breath crossed each channel. The sensor was imaged with a 12-bit digital camera and a separate ultraviolet (UV) camera for changes in the red, green, and blue color spectra, as well as 4 points in the UV spectra. Images were taken at the time the sensor was inserted into the nest, after the sensor was heated to 30°C, at the start of breath collection, after every 1 L of breath, and then every minute for 10 minutes during sensor flushing with room air (intended to help liberate VOCs trapped in the oxidizer and desiccant).

Statistical Methods

Color changes were converted to numerical vectors for change in the red, green, blue, and four UV spectra. The statistical prediction model-building procedure included four steps. The first step was feature extraction. Features were derived from the images taken during patient breathing, and then separately from the images taken both during breathing and the filtered room air flush. The features derived from the images taken during patient breathing included the mean and standard deviation of the change from baseline as well

Figure 1. Colorimetric sensor array cartridge. The chemoresponsive elements include metalloporphyrins, base indicators, acid indicators, redox dyes, solvatochromic dyes, and nucleophilic indicators. There are 4 channels for airflow, each containing 32 sensor elements. The *top channel* contains a desiccant and oxidizer and the *bottom channel* contains an oxidizer.

as the mean and standard deviation of the log-transformed raw values. To derive features from the images taken both during breathing and the filtered room air flush a nonparametric local polynomial regression as well as a simple linear regression were produced from the data for each color time series. Four model-based features were derived for each time series: the area under the curve for nonparametric regression, residual standard error of the fitted curve, total variation of the fitted curve, and linear growth trend of the data.

The second step was feature filtering for variable dimension reduction. A univariate logistic regression was fit for each feature. Features whose C-statistic was greater than 0.6 were identified as potential predictor variables. The third step was model training. Our data set was randomly split into a training set and a testing set at a 7:3 ratio. A variable selection procedure and correlation analysis were conducted to avoid multicollinearity and overfit in the model. Random forest models were built, using the subset of variables selected from the training set, with and without the inclusion of clinical variables known to be risk factors for the development of lung cancer (age, smoking history, and COPD). The fourth step was model validation. The fitted random forest models were evaluated on the testing set. To avoid randomness in data split, we repeated the third and fourth steps 100 times and summarized the prediction accuracy results. For comparison, models were built in a similar fashion using only the clinical features, and separately using all study subjects (rather than the 7:3 training: testing split).

In exploratory analyses we normalized the sensor changes to the patient's end-tidal CO₂ values. A ratio of the mean of the peak end-tidal CO₂ of an individual subject (based on the first five breaths) to the mean value of all subjects was developed for CC subjects when available. The sensor color changes were divided by this ratio to normalize for ventilation efficiency. To determine the impact of variation in VOC profiles related to demographics and comorbidities we performed two additional analyses. First, where the ratio of subjects with cancer to control subjects exceeded 1:3 we performed propensity matching using clinical variables to obtain a more homogeneous group with a ratio of 1:2 and repeated the analysis. Second, we selected the largest clinical phenotype based on age

Table 1. Study population

	Subjects with Cancer (n = 97)	Control Subjects (n = 182)	Total (n = 279)	P Value
Age (yr), mean (SD)	66.1 (9.7)	63.0 (9.1)	64.1 (9.4)	0.0068
Smoking (pack-years), mean (SD)	45.0 (28.1)	40.7 (24.6)	42.2 (25.9)	0.2387
Sex				0.4125
Female	46 (47.4%)	77 (42.3%)	123 (44.1%)	
Male	51 (52.6%)	105 (57.7%)	156 (55.9%)	
Smoking history				0.5535
Current	32 (33.0%)	50 (27.5%)	82 (29.4%)	
Former	56 (57.7%)	110 (60.4%)	166 (59.5%)	
Never	9 (9.3%)	22 (12.1%)	31 (11.1%)	
COPD	33 (34.0%)	73 (40.1%)	106 (38.0%)	0.3183
DM	9 (9.3%)	23 (12.6%)	32 (11.5%)	0.4017
Elevated cholesterol	36 (37.1%)	63 (34.6%)	99 (35.5%)	0.6779

Definition of abbreviations: COPD = chronic obstructive pulmonary disease; DM = diabetes mellitus.

(<55, 55–70, >70 yr), sex, and the presence of COPD, and repeated the analysis within this group.

Demographic variables were described using sample mean with standard deviation or proportion as appropriate. Categorical variables were compared using the Pearson chi-square test, and continuous variables were compared using the two-sample independent *t* test. Accuracies are reported as C-statistics, which represent the area under a receiver operating characteristic curve, where 1.0 represents a perfect test and 0.5 represents chance performance. All analyses were performed with the R statistical package (www.r-project.org).

Results

Two hundred and seventy-nine subjects were enrolled, 97 with lung cancer (92 CC) and 182 control subjects (107 CC). Subjects with cancer were slightly older than control subjects (66.1 vs. 63.0 yr; *P* = 0.0068). There were no differences in other demographic variables or relevant comorbidities (Table 1). The control group included 126

at-risk subjects and 56 who presented with indeterminate lung nodules. The at-risk subjects were less likely to be never-smokers than the nodule control subjects (1.6 vs. 35.7%; *P* < 0.0001). All other demographics and relevant comorbidities were similar. The mean nodule diameter was 11.8 mm. Of the 97 subjects with lung cancer, 7 were small cell, 56 adenocarcinoma, 30 squamous cell carcinoma, 1 large cell, and 3 non-small cell NOS. There was a nearly equal distribution of localized and advanced stages of lung cancer (Table 2).

Models comparing cancer and histology subgroups with control subjects were developed and tested. The accuracies of the models when applied to the test sets, reported as C-statistics, ranged from 0.794 to 0.861. Models built from the entire data set were marginally more accurate (C-statistic, 0.800–0.864). The accuracies were higher when the histology subgroups were compared with control subjects. There was little difference in the model accuracies when exhaled breath features alone were used to develop the models, compared with models that included clinical variables. Models based on clinical features alone

Table 2. Number of lung cancers of each histologic subtype and stage

	Stage I	Stage II	Stage III	Stage IV	Total
Adenocarcinoma	19	8	14	15	56
Squamous	13	4	11	2	30
Other NSCLC	1	0	0	3	4
Small cell	1	0	3	3	7
Total	34	12	28	23	97

Definition of abbreviation: NSCLC = non-small cell lung cancer.

Table 3. Accuracy of models*

Models (no. of subjects)	Trained on 70%, Tested on 30%				Trained and Tested on All: Breath Only
	Breath Only [†]	Breath + Clinical [†]	Clinical Only [†]	Likelihood Ratio	
All cancer vs. control	0.794 (0.775–0.819)	0.802 (0.778–0.822)	0.495 (0.479–0.510)	3.6	0.803
NSCLC vs. control	0.810 (0.791–0.823)	0.812 (0.793–0.825)	0.479 (0.472–0.496)	3.6	0.813
Adenocarcinoma vs. control	0.826 (0.813–0.838)	0.817 (0.803–0.830)	0.519 (0.506–0.532)	5.2	0.802
Squamous vs. control	0.854 (0.834–0.872)	0.861 (0.841–0.880)	0.451 (0.434–0.468)	8.8	0.864
Stage I vs. control	0.854 (0.840–0.867)	0.852 (0.838–0.867)	0.477 (0.447–0.507)	10.4	0.838
Stage I adenocarcinoma vs. control	0.868 (0.851–0.885)	0.867 (0.849–0.884)	0.481 (0.464–0.498)	14.9	0.880
Stage I squamous vs. control	0.896 (0.874–0.918)	0.896 (0.875–0.917)	0.514 (0.473–0.554)	9.9	0.842
Adenocarcinoma vs. squamous	0.881 (0.861–0.901)	0.887 (0.868–0.906)	0.577 (0.556–0.597)	2.4	0.908
Stage I vs. IV	0.937 (0.918–0.955)	0.948 (0.937–0.960)	0.544 (0.513–0.575)	4.9	0.955

Definition of abbreviation: NSCLC = non-small cell lung cancer.

*Validated C-statistics (with 95% confidence intervals) through model training on 70% of subjects and testing on 30% appear in the second through fourth columns. C-statistic of model built from entire data set appears in the sixth column.

[†]Breath Only = models using features from the breath test only; Breath + Clinical = models using features from the breath test and clinical variables (age, smoking history, chronic obstructive pulmonary disease [COPD]); Clinical Only = models using clinical features only (sex, smoking history, COPD).

were much less accurate than models based on exhaled breath results (Table 3).

The model accuracies of stage I cancer versus control were equally, or more, accurate although the numbers of subjects at stage I were relatively small. Models developed and tested to characterize cancer histology, and to compare early- with late-stage cancer, were accurate. There was little difference in accuracy when CC and NJH subjects were included compared with when CC subjects alone were included. Normalization of sensor changes to end-tidal CO₂ values, where available, did not influence the accuracy of the model of all cancers versus control subjects (C-statistics, 0.784 for not normalized vs. 0.776 for normalized).

In the analyses performed to assess the influence of the subjects' clinical

phenotypes, propensity matching the cancers and control subjects for clinical variables led to models with improved accuracy. The model developed using only the largest available subgroup (male, age 55–70 yr, without COPD), containing 49 subjects (18 cancer, 31 control), was accurate (Table 4).

Discussion

Lung cancer is a disease in need of accurate molecular biomarkers. Here we report the results of the development of a CSA-based profile of exhaled breath VOCs as a biomarker that could assist with the diagnosis and characterization of lung cancer. We found that the CSA profile had good accuracy at separating subjects with

lung cancer from clinically relevant control subjects; that the accuracy improved when subtypes of lung cancer were compared with control subjects; and that the accuracy was highest when the signatures were developed within a specific subset of subjects defined by their clinical phenotype. Finally, the results showed promise at being able to characterize the lung cancer's histology and stage.

The current report extends our prior research, in which we used older versions of the CSA (14, 15). The CSA used in this study has advanced in many ways (18). The sensor elements are now pigments contained in ormosils that provide more surface area for reaction than prior dye-based formulations. The number of pigments available has increased; a peroxidation phase was added to two of

Table 4. Accuracy of models matching for clinical phenotype*

	All Subjects [†]	Matched [†]	Clinical Only [†]	Likelihood Ratio, Matched
Adenocarcinoma vs. control	0.826 (0.813–0.838)	0.866 (0.853–0.878)	0.528 (0.508–0.549)	5.3
Squamous vs. control	0.861 (0.841–0.880)	0.882 (0.866–0.898)	0.567 (0.529–0.605)	6.2
Stage I vs. control	0.854 (0.840–0.867)	0.896 (0.878–0.913)	0.599 (0.556–0.642)	6.0
Stage I adenocarcinoma vs. control	0.868 (0.851–0.885)	0.903 (0.880–0.926)	0.597 (0.561–0.634)	5.5
Stage I squamous vs. control	0.896 (0.847–0.918)	0.959 (0.939–0.976)	0.677 (0.631–0.723)	14.4
All cancer vs. control (M, age 55–70 yr, no COPD)	0.982 (0.980–0.984)			27.6

Definition of abbreviations: COPD = chronic obstructive pulmonary disease; M = male.

*Validated C-statistics through model training on 70% of subjects and testing on 30%.

[†]All Subjects = best model built with data from all study subjects within the listed groups (see Table 3); Matched = best model built after propensity matching each subject with cancer in the listed groups to two control subjects; Clinical Only = models using clinical features only (age, smoking history, COPD).

the channels; the imaging system has become more sophisticated, including imaging in the UV spectra; and the breath collection instrument has advanced to include VOC filtering of inspired air and end-tidal CO₂ control of alveolar breath sampling. The models developed using the results of exhaled breath testing only (i.e., not including clinical variables) in the current study outperformed those from our prior work despite a more conservative statistical method for assessing their accuracy in this study (e.g., C-statistic of model developed for non-small cell carcinoma vs. control in current study was 0.810, compared with 0.701 in the prior study).

Lung cancer is a heterogeneous disease that occurs in a diverse group of people. Lung cancer can be indolent or aggressive; is histologically and molecularly variable; occurs in men and women, smokers and nonsmokers, and people spanning a wide range of ages and with a variety of comorbidities. The unique metabolic background of an individual could influence the accuracy of a single metabolic biomarker. A patient's clinical phenotype could be useful as a baseline to help distinguish a noncancer from a cancer metabolic biosignature. Matching a metabolic biosignature to a library of signatures may allow the most accurate identification of underlying lung cancer. Our results support improved accuracy of our metabolic biosignature when developed within a relatively uniform clinical phenotype.

There are many influences on the composition of exhaled breath VOCs. The optimal means of collecting and assessing them remains unclear. We chose to filter ambient VOCs, have our subjects perform tidal breathing, use end-tidal CO₂ monitoring to identify the alveolar portion of the breath, expose the sensor to the breath in real time, and request that no food or drink be consumed for at least 1 hour before the test. In an exploratory analysis, normalization of sensor responses to the end-tidal CO₂ value did not appear to influence the accuracy of the results. As a step toward technical validation of this methodology each of these components of breath analysis should be assessed for their impact on the robustness and consistency of the signal obtained. Clinical validation will require that a larger cohort be studied, using a technically validated system, after developing and fixing a cancer signature or signatures based on clinical phenotype.

The CSA is a cross-responsive sensor. This means the signature obtained reflects the entire mixture of VOCs to which the sensor was exposed. This signature cannot tell us what the individual constituents of that mixture are. Other technologies, such as gas chromatography-mass spectrometry, have been used in efforts to define the individual VOCs that make up the mixture (19). A unifying model has not emerged, in part due to limitations of the technologies, variability in the populations studied, differences in breath sampling techniques, and the heterogeneity of lung cancer. Further research with gas chromatography-

mass spectrometry and similar technologies will help to elucidate the pathophysiologic basis of the VOC signatures (20). The ease of use and low cost of sensor technologies, such as the CSA, make them more apt to be used in the clinical setting.

Other limitations of our study include the small sample size for some of the comparisons where the accuracy was highest. These comparisons should be viewed as exploratory, helping to guide the next phase of breath biomarker development. In addition, it is not clear that the breath collection methodologies used in this study are optimal, and minor inconsistencies in sensor manufacturing could impart unseen biases in the results. These issues will need to be addressed as part of the validation of this biomarker for clinical use.

The accuracies reported here compare favorably with other available biomarkers of early detection and/or nodule management. Breath analysis may be the least invasive biomarker source, and the only one that could lead to a near real-time result. As such, these results support further research in this field.

In conclusion, the CSA signature of exhaled breath VOCs was capable of distinguishing patients with lung cancer from clinically relevant control subjects in a discovery-level trial. The incorporation of clinical phenotypes into the further development of this biomarker may optimize its accuracy. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

References

- Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, Majem M, Lopez-Vivanco G, Isla D, Provencio M, *et al.*; Spanish Lung Cancer Group. Screening for epidermal growth factor receptor mutations in lung cancer. *N Engl J Med* 2009;361:958–967.
- Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou SH, Dezube BJ, Jänne PA, Costa DB, *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 2010;363:1693–1703.
- Chapman CJ, Healey GF, Murray A, Boyle P, Robertson C, Peek LJ, Allen J, Thorpe AJ, Hamilton-Fairley G, Parsy-Kowalska CB, *et al.* EarlyCDT®-Lung test: improved clinical utility through additional autoantibody assays. *Tumour Biol* 2012;33:1319–1326.
- Li XJ, Hayward C, Fong PY, Dominguez M, Hunsucker SW, Lee LW, McLean M, Law S, Butler H, Schirm M, *et al.* A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. *Sci Transl Med* 2013;5:207ra142.
- Sozzi G, Boeri M, Rossi M, Verri C, Suatoni P, Bravi F, Roz L, Conte D, Grassi M, Sverzellati N, *et al.* Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. *J Clin Oncol* 2014;32:768–773.
- Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007;13:361–366.
- Weinberg F, Hamanaka R, Wheaton WW, Weinberg S, Joseph J, Lopez M, Kalyanaraman B, Mutlu GM, Budinger GRS, Chandel NS. Mitochondrial metabolism and ROS generation are essential for Kras-mediated tumorigenicity. *Proc Natl Acad Sci USA* 2010;107:8788–8793.
- Martin RCG, Li Y, Liu Q, Barker DF, Doll MA, Hein DW. Manganese superoxide dismutase expression as a function of genotype and lung cancer pathology. *Cancer Invest* 2010;28:813–819.
- Park JY, Matsuo K, Suzuki T, Ito H, Hosono S, Kawase T, Watanabe M, Oze I, Hida T, Yatabe Y, *et al.* Impact of smoking on lung cancer risk is stronger in those with the homozygous aldehyde dehydrogenase 2 null allele in a Japanese population. *Carcinogenesis* 2010;31:660–665.
- An Z, Chen Y, Zhang R, Song Y, Sun J, He J, Bai J, Dong L, Zhan Q, Abliz Z. Integrated ionization approach for RRLC-MS/MS-based metabolomics: finding potential biomarkers for lung cancer. *J Proteome Res* 2010;9:4071–4081.
- Sponring A, Filipiak W, Ager C, Schubert J, Miekisch W, Amann A, Troppmair J. Analysis of volatile organic compounds (VOCs) in the

- headspace of NCI-H1666 lung cancer cells. *Cancer Biomark* 2010;7: 153–161.
- 12 Poli D, Goldoni M, Corradi M, Acampa O, Carbognani P, Internullo E, Casalini A, Mutti A. Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatization SPME-GC/MS. *J Chromatogr B Analyt Technol Biomed Life Sci* 2010;878: 2643–2651.
 - 13 Peng G, Tisch U, Adams O, Hakim M, Shehada N, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, Haick H. Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nat Nanotechnol* 2009;4:669–673.
 - 14 Mazzone PJ, Hammel J, Dweik RA, Na J, Czich C, Laskowski D, Mekhail T. Lung cancer diagnosis by the analysis of exhaled breath with a colorimetric sensor array. *Thorax* 2007;62:565–568.
 - 15 Mazzone PJ, Wang XF, Xu Y, Mekhail T, Beukemann MC, Na J, Kemling JW, Suslick KS, Sasidhar M. Exhaled breath analysis with a colorimetric sensor array for the identification and characterization of lung cancer. *J Thorac Oncol* 2012;7:137–142.
 - 16 Mazzone PJ, Wang X, Rhodes P, Martino R, Lim S, Beukemann M, Seeley M, Choi H, Jett J. The analysis of volatile organic compound profiles in the breath as a biomarker of lung cancer [abstract]. *Chest* 2013;144(4_MeetingAbstracts):645A.
 - 17 Gould MK, Donington J, Lynch WR, Mazzone PJ, Midhun DE, Naidich DP, Wiener RS. Evaluation of individuals with pulmonary nodules: When is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013;143(5 Suppl):e93S–e120S.
 - 18 Lim SH, Feng L, Kemling JW, Musto CJ, Suslick KS. An optoelectronic nose for the detection of toxic gases. *Nat Chem* 2009;1:562–567.
 - 19 Haick H, Broza YY, Mochalski P, Ruzsanyi V, Amann A. Assessment, origin, and implementation of breath volatile cancer markers. *Chem Soc Rev* 2014;43:1423–1449.
 - 20 Hakim M, Broza YY, Barash O, Peled N, Phillips M, Amann A, Haick H. Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem Rev* 2012;112:5949–5966.