Urinary Volatile Compounds as Biomarkers for Lung Cancer

Yosuke HANAI,1 Ken SHIMONO,1 Koichi MATSUMURA,2 Anil VACHANI,3
Steven ALBELDA,3 Kunio YAMAZAKI,2 Gary K. BEAUCHAMP,2 and Hiroaki OKA1,†

1Bioscience Technology Development Office, Panasonic Corporation, Kyoto 619-0237, Japan
2Monell Chemical Senses Center, Philadelphia, PA 19104-3308, USA
3University of Pennsylvania Medical Center, Philadelphia, PA 19104-4318, USA

Received October 12, 2011; Accepted December 27, 2011; Online Publication, April 7, 2012
[doi:10.1271/bbb.110760]  

Lung cancer is a leading cause of deaths in cancer. Hence, developing early-stage diagnostic tests that are non-invasive, highly sensitive, and specific is crucial. In this study, we investigated to determine whether biomarkers derived from urinary volatile organic compounds (VOCs) can be used to discriminate between lung cancer patients and normal control patients. The VOCs were extracted from the headspace by solid-phase microextraction and were analyzed by gas chromatography time-of-flight mass spectrometry. Nine putative biomarkers were identified as elevated in the lung cancer group. Receiver operating characteristic curve analysis was also performed, and the markers were found to be highly sensitive and specific. Next we used principal component analysis (PCA) modeling to make comparisons compare within the lung cancer group, and found that 2-pentanone may have utility in differentiating between adenocarcinoma and squamous cell carcinomas.

Key words: lung cancer; volatile organic compounds (VOCs); gas chromatography with time of flight mass spectrometry (GC-TOF MS); biomarker; urine

The WHO estimates that 17 percent of all cancer related deaths are due to lung cancer people1,2 and annually lung cancer is thought to be responsible for the death of about 1.3 million worldwide. Given the profound burden of disease due to lung cancer, novel diagnostic strategies with the potential to reduce lung cancer mortality are required, especially ones with a capacity to allow for early detection of lung cancer and thus facilitate improvements in prognosis. Imaging of high-risk patients is emerging as the dominant approach to early diagnosis, although large-cohort studies to validate this approach are ongoing.2–4) While imaging modalities offer enhanced sensitivity, they may lack the necessary specificity. Recent studies using CT have shown that about 5.26% of high-risk smoking patients have detectable lung nodules, but only about 4% (range 2–11%) of these nodules are malignant.5) Clearly surgical resection of all of these nodules is neither practical nor desirable, and hence approaches that allow clinicians to judge better which nodules should be removed are needed. One promising strategy is to combine imaging methodologies with lung cancer biomarkers as a mechanism to increase specificity.6–8) Since the incidence of lung cancer in this “nodule population” is significantly higher than in current or former smoking populations, if applied in this context, the biomarkers do not need to possess the high sensitivities and specificities required for much larger population-based screening. These types of biomarkers may also have clinical utility in following up patients post-treatment. Current biomarker candidates from the blood, sputum, and urine consist of many classes of molecules, including proteins, tumor antigens, anti-tumor antibodies, cell type-specific peptides, various metabolic products, and epigenetic phenomena such as hyper-methylated DNA, RNA, and specific gene expression,9) but to date none of the biomarkers identified possess the required sensitivity, specificity, and reproducibility necessary for application as a non-invasive method for the detection and monitoring of lung cancer growth. Recently, metabolomics has developed rapidly; it is used to obtain information about the cellular processes of an organism. The metabolites might reflect physiological functions and pathological characteristics in more detail, because the metabolome is the endpoint of the omics cascade. Cancer cells require relatively large amounts of energy for growth, and the mechanisms of energy production in many cancer cells are different from that in normal cells.10) Cancer cells use large amounts of glucose and glutamine as energy sources and are dependent on glycolysis rather than oxidative phosphorylation for energy production, even in the presence of sufficient oxygen.10) This suggests that the presence of a tumor leads to changes in the levels of low molecular weight compounds in the TCA cycle and glycolysis.

Small molecular weight volatile organic compounds (VOCs) are a class of biomarkers thought to have potential for the detection of lung cancer. Studies have shown that lung cancer cell lines release specific VOCs in vitro.11) For example, a recent study using solid phase micro-extraction followed by gas chromatography found that the 1-butanol and 3-hydroxy-2-butanol concentrations were significantly higher than in the breath of lung cancer patients compared with controls.12) The “volatile hypothesis” for lung cancer has led to a number of

1 To whom correspondence should be addressed. Fax: +81-774-98-2572; E-mail: oka.hiroaki@jp.panasonic.com

Abbreviations: VOCs, volatile organic compounds; TMN, tumor node metastasis; AJCC, American Joint Committee on Cancer; SPME, solid-phase microextraction; DVB, divinylbenzene; CAR, carboxen; PDMS, polydimethylsiloxane; GC-TOF MS, gas chromatography with time of flight mass spectrometry; AUC, area under the curve; ROC, receiver operating characteristic; PCA, principal components analysis; BER, balanced error rate
studies examining the utility of analyzing these compounds in exhaled breath using animals (such as dogs)\textsuperscript{13} or sophisticated biochemical techniques,\textsuperscript{14,15} but collecting, condensing, storing, and analyzing exhaled breath samples is problematic, and hence a more convenient source of volatile compounds, such as those detected in urine samples, is preferable. Matsumura \textit{et al.} trained mice to discriminate an odorant of urine derived from mice with implanted tumors.\textsuperscript{16} Based on these considerations, observation of changes in the VOCs in the urine might be used to detect lung cancer. In the present study, we carried out GC-TOF MS-based metabolite profiling of lung cancer using urine obtained from lung cancer patients, and investigated to determine whether the presence of a specific pattern of VOCs distinguishes lung cancer patients from healthy controls.

**Materials and Methods**

**Subject and urine preparation.** The study included two groups: patients with lung cancer and normal volunteers. We enrolled 20 patients and 20 normal controls from the University of Pennsylvania Medical Center. The characteristics of the two patient groups are described in Table 1A. A diagnosis of lung cancer was confirmed in each patient by bronchoscopy and biopsy findings. A given tumor was staged using the tumor, node, metastasis (TNM) system for lung cancer according to the AJCC (American Joint Committee on Cancer) Cancer Staging Manual, 6th Edition.\textsuperscript{17} The histological diagnoses and corresponding lung cancer states are described in Table 1B. Diagnoses of the normal volunteers are summarized in Table 1C. Immediately after collection, urine samples were stored at \(-80^\circ\text{C}\) until use. To obtain low-molecular-weight compounds (<300 Da), the urine sample was centrifuged at 13,000 \(g\) for 10 min, and then the supernatant was filtered at 4°C using 30-kDa, 10-kDa, and 3-kDa membranes (Amicon YM-30, YM-10, and YM-3 Milipore, Bedford, MA, USA).

**Extraction of urinary volatile compounds by solid-phase extraction.** To select a suitable solid-phase-microextraction (SPME) fiber, we sampled the volatile compounds using four types of SPME fiber (CAR/PDMS, DVB/CAR/PDMS, PDMS/DVB, Polyacrylate). On the basis of our findings, the 2 cm 50/30 \(\mu\text{m}\) DVB/CAR/PDMS (divinylbenzene/carboxen/polydimethylsiloxane; Supelco, Bellefonte, PA) fiber was selected. Then we chemically analyzed 40 urine samples (20 patients, 20 controls). Urinary volatile compounds were extracted by head space SPME methods (HS-SPME) using an auto sampler (Combi-pal, CTC Analytics). The VOCs of the urine were extracted by head space SPME methods (HS-SPME) using a Combi-pal auto sampler (CTC Analytics). A 100-\(\mu\text{L}\) sample of the urine was applied to a 2-mL crimp top vial sealed by magnetic crimp cap. The vial was set on the Combi-pal and equilibrated for 10 min at 45°C. The volatile compounds in the headspace were extracted by SPME fiber for 50 min at 45°C.

**Gas chromatography and mass spectrometry.** Gas chromatography with a time of flight mass spectrometry (GC-TOF MS) system was used with of 7890a GC (Agilent, Palo Alto, CA) equipped with an auto sampler (Combi-pal, CTC Analytics, Zwingen, Switzerland) and a GCT Premier (Waters, Milford, MA) for time of flight mass spectrometry. The SPME fiber with absorbed volatile compounds was inserted into the injection port of the 7890a GC using the auto sampler, and desorbed for 10 min at 240°C. The injection was pulse splitless (closed 3 min) with a 0.75-mm liner. The GC-TOF MS system was equipped with an InertCap 5 T.L. column (60 m + 2 m transfer line, 0.25 mm i.d., 0.25 \(\mu\text{m}\) thick; GL science, Tokyo, Japan), which was used for separation and analysis of the desorbed volatiles. We employed the following chromatographic protocol for separation before MS analyses: 40°C for 5 min, then 5°C/min to 240°C with a 5-min hold at the final temperature. The column flow was constant at 1 mL/min. The injection port was held at 240°C. Operating parameters for the mass spectrometer were as follows: ion source temperature, 200°C; ionization energy, 70 eV; scanning frequency 0.2 s/spectrum from \(m/z\) 40 to \(m/z\) 500. Peak identification was accomplished by manual interpretation of the spectra and matching against a mass spectral library (NIST/EPA/NIH mass spectral library, NIST 08; Mass Spectral Library of Drugs, Poisons, Pesticides, Pollutants, and Their Metabolites, Wiley) and comparison with commercially available standard samples when available.

**Data processing and quantitative analysis.** The chromatographic peak areas were integrated using MassLynx 4.1 (Waters, Milford, MA). Genesis peak detection was applied to create a list of putatively annotated ions. Detection and integration of the generated ion peaks from electron ionization was done using XCMS software package version 1.16.3 (http://massspec.scripps.edu),\textsuperscript{18} running under R version 2.10.1 (http://cran.r-project.org/). The significance levels of differences between groups were calculated by Student's \(t\)-test. We also determined that the increased extracted ion peaks had a \(p\)-value of \(<0.05, 1.5\) times higher than the averaged peak area of the other group. Of the peaks, a manual inspection of the EIC for each peak was made to validate the detected peak. Refinement of VOC was done by manual deconvolution based on the retention times and peak shapes of the various increased extracted ion peaks.

The VOCs identified were quantified using commercially available reagents. For each of the reagents, stock solutions were made to a concentration of 100 nm (Sigma-Aldrich, St. Louis, MO) by dissolving them into a 1-mL mixture of water and methanol (1/1 v/v). Calibration solutions of 10, 50, 100, 500, 1000, and 10,000 nm were also made up. Standard curves were created based on the peak areas, which were obtained from HS-SPME GC-TOF MS analysis of the calibration solution. The data were analyzed in triplicate. \(p\)-values were obtained by Student's \(t\)-test.

**Statistical analyses.** Differentiation performance (specificity and sensitivity) was assessed using the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) curves. The ROC curves,
curves and PCA were also employed. Squamous cell carcinoma and the adenocarcinoma samples, ROC analysis package. To investigate metabolomic differences between the normal controls, principal components analysis (PCA) was employed using the package ‘prcomp,’ which is a part of the R statistical software. Nine out of the 19 VOCs had high similarity to the deconvolution analysis using retention time of each peak as an indicator, 19 VOCs were isolated of the database, and were identified by matching retention time, m/z value, intensity, p-value, and fold change. The chromatogram of the urine a given single person contained about 940 ion peaks. The ion peak with a p-value <0.05 and an area 1.5-fold greater than the averaged peak area of the controls was defined as an increased extracted ion peak for patients with lung cancer. From each EIC, 101 increased extracted ion peaks were selected for further investigation. As a result of the deconvolution analysis using retention time of each peak as an indicator, 19 VOCs were isolated. A similarity search of the 19 VOCs was done using XCMS. XCMS software was used for peak-matching, non-linear retention time alignment, and quantitation of mass spectral ion intensities. By aligning, we obtained information on each ion peaks as to discrimination ability of the various volatile organic compounds.

### Results

#### Comparative analysis of VOCs from the urine of lung cancer patients and normal controls

Urine samples (100 μL) from 20 lung cancer patients and 20 normal control volunteers were analyzed by HS-SPME GC-TOF MS. The TIC data were difficult to analyze because they contained many peaks, and hence it was difficult to distinguish between the peaks from the lung cancer and the control groups. To detect increased ion peaks from the lung cancer group, comparative analysis of the lung cancer and the control group was done using XCMS. XCMS software was used for peak-matching, non-linear retention time alignment, and quantitation of mass spectral ion intensities. By alignment, we obtained information on each ion peaks as to retention time, m/z value, intensity, p-value, and fold change. The chromatogram of the urine a given single person contained about 940 ion peaks. The ion peak with a p-value <0.05 and an area 1.5-fold greater than the averaged peak area of the controls was defined as an increased extracted ion peak for patients with lung cancer. From each EIC, 101 increased extracted ion peaks were selected for further investigation. As a result of the deconvolution analysis using retention time of each peak as an indicator, 19 VOCs were isolated (Table 2).

#### Quantitative analysis of increased VOCs in the urine of patients with lung cancer

To determine the concentrations of the VOCs, a standard curve was prepared by analysis by HS-SPME. Standard curves were induced using the peak area at the most abundant m/z value. The standard curve of each VOC indicated good linearity in a range of 10 nM to 10,000 nM (R² = 0.99). The urinary concentration of each VOC was then calculated using the respective standard curve. The quantitative findings, with accompanying statistical analysis, are summarized in Table 4. The nine VOCs were significantly elevated in the lung cancer group relative to the control group (p < 0.05). In particular, the six VOCs, tetrahydrofuran, 2-chloroethanol, cyclohexanone, 2-ethyl-1-hexanol, 2-phenyl-2-propanol, and isophorone, were significantly different from the control group (p < 0.01).

#### Discrimination ability of the various volatile organic compounds

Receiver Operating Characteristic (ROC) curve analysis was used to assess whether the urine derived the VOCs could be used to differentiate between the lung cancer patients and the normal controls (Table 5). The cutoff value for each compound was calculated by minimizing the Balanced Error Rate (BER). The following compounds were found to have 95% sensi-
A principal components analysis (PCA) model was constructed using the concentrations of the nine VOCs as variables. The PCA score plots of the various samples were scattered into two different regions, with the exception of one normal control (Fig. 1). PC1 was found to be a principal factor in this separation. It indicated the contribution of all nine VOCs. To validate the robustness of the PCA model in discriminating the lung cancer patients from the normal controls, ROC analysis was performed to evaluate the urinary concentrations of 2-pentanone. Hence ROC analysis was performed to evaluate the urinary concentrations of 2-pentanone further. The AUC value, sensitivity, and specificity were found to be 0.766, 100%, and 70% respectively. This finding provides evidence that urinary 2-pentanone has utility in distinguishing between adenocarcinomas and squamous cell carcinomas. The patients diagnosed with adenocarcinoma were concentrated in a narrow range, whereas these diagnosed with squamous cell carcinoma were scattered across a wide range according to the PCA score field (Fig. 2). In addition, the patients with adenocarcinoma tended to have higher concentrations of 2-pentanone. Hence ROC analysis was performed to evaluate the urinary concentrations of 2-pentanone further. The AUC value, sensitivity, and specificity were found to be 0.766, 100%, and 70% respectively. This finding provides evidence that urinary 2-pentanone has utility in distinguishing between adenocarcinomas and squamous cell carcinomas.

Discussion

Conventional tumor markers are high molecular compounds such as fragmented proteins, hormone, enzymes, and isozymes. The process of detecting these markers is invasive, typically requiring analysis of blood products. Hence an ability to use less invasive diagnostic methods to aid in the diagnosis of cancer in patient populations is a welcome development. In this study, we focused on urinary VOCs, and found that...
through urinary VOCs, one can differentiate between normal controls and patients with lung cancer. We identified nine VOCs, which were found to be present at higher concentrations in the urine of patients with lung cancer than in that of the normal controls.

We also used SPME analysis. The advantage of SPME is that extraction is simple and fast and can be done without solvents. To select the best SPME fiber for use in human urine, we tested four kinds of fibers (CAR/PDMS, DVB/CAR/PDMS, PDMS/DVB, Polyacrylate). The DVB/CAR/PDMS fiber induced the most VOCs, DVB/CAR/PDMS, PDMS/DVB, Polyacrylate, while the CAR and PDMS fibers were measured at high temperature because of their high water solubility. Additionally, Westhoff et al. have reported that comparing exhaled breath of lung cancer patients and healthy subjects, branched hydrocarbons and alkanes were more abundant in the lung cancer patients.21) Deng et al. also reported using volatile biomarkers in the blood of patients with lung cancer.20) In the present study, we focused on VOCs obtained through the urine of lung cancer patients. We found that a number of the compounds we identified the including ketones (2-pentanone, cyclohexanone, and isophorone) and alcohols (2-ethyl-1-hexanol and 2-phenyl-1-propanol) were different from previous reports. The reasons might be as following: (i) individual differences in the patients; (ii) differences in the VOC extraction methods used; (iii) differences in the source of VOCs (urine, breath, or blood).

In this study, the urine samples were collected from a single medical institution, and hence it is possible that there were factors common to these patients, such as food or medication. This raises the further possibility that some of the VOCs identified were derived from exogenous sources such as food or medicine hence might not be related to lung cancer. For instance, one of the identified VOCs, 2,6-diisopropylphenol, is a principal component of propofol, a medicine used to induce general anesthesia. 2-Chloroethanol is derived from ethylene oxide and is used in the sterilization of medical devices.23) 2-Phenyl-2-propanol is a metabolite of cumene and of phenobarbital.24,25) Phenobarbital is an anxiolytic drug, and is likely to be prescribed to alleviate the anxiety of patients with lung cancer. These exogenous compounds, associated with anxiolytics, anesthetics, and sterilizations, which might be associated particularly with patients with lung cancer, are not considered suitable as biomarker. Isophorone (3,5,5-trimethyl-2-cyclohexene-1-one) is colorless liquid having a peppermint odor, widely used in industry as a solvent of natural and synthetic resins, waxes, oils, pesticides, paints, and printing inks.27) Although it has been detected in human urine,38,39) it might be a characteristic contaminant in patients. Cyclohexanone might be an oxidative product of cyclohexane. Cyclohexane was found by Phillips et al. to be a breath biomarker for lung cancer patients.22) A possible reason that cyclohexanone was found in the urine is that it might have been excreted in the urine due to increased water solubility. Additionally, Westhoff et al. reported that healthy persons and COPD patients (both with and without lung cancer) were distinguished by the concentration of cyclohexanone in breath.40) But cyclohexanone has been detected in human urine as a contaminant derived from administration sets.41,42) Since both iso-
phorone and cyclohexanone are likely to be caused by smoking. However, both the lung cancer patients and the healthy controls were similarly exposed to smoke of tobacco. Moreover, smoking and carcinogenic rates are proportional. Hence compounds derived from smoke of tobacco are considered useful as biomarker for initial screening.

The biological significance of the VOCs identified as being biomarkers for lung cancer is not yet clear, but there is some evidence that the VOCs identified are related to lung cancer. In a study of cultured cells, 2-methylpentane, 2-ethyl-1-hexanol and 2-pentanone were found to be more abundant in the headspace of NCI-H2087 lung cancer cell line and A549 lung cancer cell line. Hence urinary VOCs 2-ethyl-1-hexanol and 2-pentanone might be products of lung tumor cells.

The collection and analysis of exhaled breath samples are problematic, whereas urine sample is suitable as a convenient source for screening study. In the present study, we found that volatile components have the potential to act as cancer markers for patients with lung cancer. Four compound 2-ethyl-1-hexanol, 2-pentanone, tetrahydrofuran, and 2-methylpyrazine, are considered good candidate initial screening biomarkers of lung cancer. This leads to the possibility that VOCs can be further developed into a convenient and non-invasive diagnostic method for lung cancer. Future experiments using large numbers of samples are required to pursue these possibilities.

References