Non-invasive Diagnosis of Pancreatic Cancer Through Detection of Volatile Organic Compounds in Urine

Ramesh Arasaradnam, Alfian Wicaksono, Harrison O’Brien, Hemant M. Kocher, James A. Covington, Tatjana Crnogorac-Jurcevic

PII: S0016-5085(17)36347-3
DOI: 10.1053/j.gastro.2017.09.054
Reference: YGAST 61530

To appear in: Gastroenterology
Accepted Date: 5 September 2017


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Title: Non-invasive Diagnosis of Pancreatic Cancer Through Detection of Volatile Organic Compounds in Urine

Authors: Ramesh Arasaradnam¹, Alfian Wicaksono⁴, Harrison O’Brien², Hemant M Kocher³, James A Covington⁴*, Tatjana Crnogorac-Jurcevic⁵*

*Shared senior authorship

Correspondence to: Professor Ramesh Arasaradnam
Department of Gastroenterology, University Hospital Coventry & Warwick, Applied Biological Sciences, University of Coventry
Division of Surgery | UHCW NHS Trust | Clifford Bridge Road | Coventry CV2
2DX, UK

¹Department of Gastroenterology, University Hospital Coventry & Warwick, Applied Biological Sciences, University of Coventry
²Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London
³Centre for Tumour Biology, Barts Cancer Institute, Queen Mary University of London
⁴School of Engineering and Medical School, University of Warwick

Funding source: Barts Pancreas Tissue Bank is supported by PCRF.

Conflict of interest: None

Acknowledgement: Barts Pancreas Tissue Bank personnel involved in this work include Vickna Balarajah, Thomas Dowe and Amina Saad. Please see https://www.bartspancreastissuebank.org.uk/
Authors contribution: RA & TC-J designed the study, RA drafted the manuscript, AW & JC undertook analysis and RA, TC-J and JC provided interpretation, AW & JC undertook statistical analysis, JC provided technical expertise, HB & HK recruited patients and sample collection, RA, AW, HK, JC & TC-J reviewed the manuscript for intellectual content.
With its incidence approaching mortality, and with over 300,000 new cases diagnosed worldwide in 2013, pancreatic ductal adenocarcinoma (PDAC) is currently the fourth leading cause of cancer-related death and predicted to become the second by 2030 (1). More than 80% of PDAC patients are diagnosed late (2), with locally invasive and/or metastatic disease, resulting in negligible 5-year survival. Thus the quest for a simple, inexpensive and non-invasive test to detect PDAC early - whilst it is still amenable to surgical resection, continues. Detection of volatile organic compounds (VOCs) has come to the fore offering a novel approach for the detection of disease. It uses the odours that emanate from urine, breath and faeces and is akin to canine ‘sniffing’. These compounds are metabolic products and/or consequence of bacterial dysbiosis produced by the disease state (3, 4). We thus postulate that either altered cellular physiology or even alteration in the microbial milieu in patients with PDAC will alter the individual’s metabolome profile, such that the resultant VOC patterns that are emitted provide a characteristic signature that can be detected.

Description of Technology

Urinary VOC was analysed using an Owlstone™Lonestar instrument based on ion mobility spectrometry (IMS). Patients’ information is summarised in Table 1. Midstream urine samples were collected and frozen at -80°C within 4 hours of collection for long-term storage. Prior to analysis, each sample was defrosted and heated to 40°C for 10 minutes. The air above the sample (known as the headspace) was then passed into the Lonestar unit and the ion mobility measured. Each sample typically takes five minutes to analyse (60 seconds per sample to scan, with five repeats). Once all the samples were analysed, an existing data processing pipeline with four different classifiers was applied (Sparse Logistic Regression, Random Forest, Gaussian Process Classifier and Support Vector Machine). Age was not deemed to be a confounder based on our previous published work (please see supplementary material). PDAC urine samples were detected with a sensitivity (SN) of 0.91 (CI: 0.83-0.96) and specificity (SP) of 0.83 (CI: 0.73-0.90), with an area under the curve (AUC) of 0.92 (CI:0.88-0.96), using a support vector machine algorithm – Figure 1c. The results were validated by random data splitting into the training and test set achieving similar results with AUC of 0.92 (CI: 0.85-0.98), and SN and SP of 0.90 (CI: 0.74-0.98) and 0.81 (CI: 0.63-0.93), respectively. We also compared early stage disease (I/II) with healthy individuals as well as early stage I/II with late stage
disease (III/IV) using the same analysis pipeline. These revealed a sensitivity (SN) of 0.91 (CI: 0.78-0.97) and specificity (SP) of 0.78 (CI: 0.69-0.86), with an area under the curve (AUC) of 0.89 (CI: 0.83-0.94), and SN of 0.82 (CI: 0.67-0.92), SP of 0.89 (CI: 0.75-0.97) with AUC of 0.92 (CI: 0.86-0.97), respectively, again using a support vector machine algorithm (Figures 1d and e).

**Video Description**

The Lonestar is a special type of Ion Mobility Spectrometer (IMS), based on a method called FAIMS, (Field Asymmetric Ion Mobility Spectroscopy). It has the advantage of being reproducible and far more sensitive than electronic noses but without the limitations of mass spectrometry, which requires specialised and expensive infrastructure. FAIMS is based on measuring the way that ionised molecules move in very high electric fields and due to the differences in this movement, the instrument is able to separate complex mixtures of chemicals that are found in biological samples. To improve separation, the magnitude of the electric field is raised through a series of values to create a 3D map of a urine sample. These are then analysed using our in house data processing pipeline, previously described (5). In brief, first a data compression wavelet and threshold is applied to reduce the size of the dataset (which in our case is more than 50,000 data points per sample). Once completed, important features are sought within the compressed dataset, using a Wilcoxon rank-sum test, with a 10-fold cross-validation step. Such identified features can then be used with different classification algorithms to calculate clinically relevant values e.g. AUC, SN and SP of unknown samples.

**Take Home Message**

We have already showed that urine is a useful bio-fluid for detection of early stage, resectable pancreatic cancer (6,7). In this proof of concept study, we demonstrate for the first time, utility of urine specimens to discriminate healthy individuals from PDAC patients also through detection of VOCs. Moreover, we have also shown the ability to separate healthy from early stage and early stage vs advanced disease. This was achieved using only 5 ml of urine sample and applying novel IMS technology, which offers a rapid, and more cost effective analysis compared with other gas analysis technologies. The obtained results demonstrate that urine analysis using IMS
platform shows promise as an additional non-invasive approach for identification of patients with pancreatic cancer.
### Table 1. Demographic information

<table>
<thead>
<tr>
<th></th>
<th>Healthy controls (n = 81)</th>
<th>Pancreatic Cancer (n=81)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean ± SD)</td>
<td>51.4 ± 10.6</td>
<td>64.3 ± 23.7</td>
</tr>
<tr>
<td>Sex (male) %</td>
<td>30.9</td>
<td>53.1</td>
</tr>
<tr>
<td>Stage*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>n/a</td>
<td>4</td>
</tr>
<tr>
<td>IIA</td>
<td>n/a</td>
<td>5</td>
</tr>
<tr>
<td>IIB</td>
<td>n/a</td>
<td>35</td>
</tr>
<tr>
<td>III</td>
<td>n/a</td>
<td>24</td>
</tr>
<tr>
<td>IV</td>
<td>n/a</td>
<td>12</td>
</tr>
</tbody>
</table>

*One case could not be assessed
References:

Figure 1 Legend

Figure 1a) Representative image of FAIMS output showing the plume of ion dispersion from a single urine sample of a healthy control (A.U. is arbitrary units).

Figure 1b) shows similar plume of ion dispersion from a single urine sample of a patient with PDAC. Figure 1c) Receiver Operating Characteristic (ROC) curve showing performance of VOCs in differentiating healthy individuals from PDAC with AUC of 0.92 (CI: 0.88-0.96); SN=0.91 (CI: 0.83-0.96); SP=0.83 (CI: 0.73-0.9);

Figure 1d) Healthy individuals from early stage I/II (AUC =0.89; CI: 0.83-0.94); SN=0.91 (CI: 0.78-0.97); SP=0.78 (CI: 0.69-0.86). Figure 1e) shows early stage PDAC (I/II) could also be successfully separated from advanced stage PDAC (III/IV) with AUC 0.92 (CI: 0.86-0.97); SN=0.82 (CI: 0.67-0.92); SP=0.89 (CI: 0.75-0.97).
Title: Non-invasive Diagnosis of Pancreatic Cancer Through Detection of Volatile Organic Compounds in Urine

Authors: Ramesh Arasaradnam¹, Alfian Wicaksono⁴, Harrison O’Brien², Hemant M Kocher³, James A Covington⁴*, Tatjana Crnogorac-Jurcevic²*

*Shared senior authorship

Supplementary Material

The instrument used in this study was a commercial FAIMS instrument (Lonestar, Owlstone UK), which is fitted with an ATLAS (Owlstone, UK) sampling system. It is highly sensitive (detecting ions up to parts per billion) and based on distinguishing different ions based on its physical properties – specifically its mobility; it does not identify per se the specific VOCs.

The samples were frozen at -80°C within 4 hours of collection and 5 ml aliquots transported on dry ice from Barts Cancer Institute to the School of Engineering, University of Warwick. The samples were thawed overnight at 4°C in a laboratory fridge. The vial with the urine was placed in the ATLAS sampling system for 10 minutes at 40°C to maximise VOC chemical release. After 10 minutes, clean dry air is pushed over the surface of the urine and into the Lonestar. The flow rate over the sample was 500 ml/min with an additional 1500 ml/min of makeup air (to make 2L/min) pushed into the Lonestar. Each sample was tested in triplicate.

In practice, ionised molecules are passed between two plates, where an asynchronous electric field is applied. This attracts, repels or has no effect of these ions. If an ion touches one of the plates it loses its charge, and thus only ions that exit the plates without touching either side are detected. To the asynchronous electric field, an additional fixed “compensation voltage” is added that counteracts the movement. Thus, by scanning through a range of different compensation voltages we are able to measure a large range of different mobilities.
The machine was set with a dispersion field of 0 to 100% in 51 steps and -6V to +6V compensation voltage in 512 steps, with both positive and negative fields applied to the samples. This created a dataset of 52,224 points per sample, with each sample taking less than 60 seconds to capture.

The FAIMS data was processed using a well-established pipeline, which we have developed over a number of years for this type of study [1]. In brief, we first use a pre-processing data compression step in the form of a 2D wavelet transform (using Daubechies D4 wavelets). This has the effect of improving and simplifying subsequent analysis steps. A threshold is then applied to remove data with little or no discriminatory power (based on our previous work). This was followed by a 10-fold cross-validation, using 80% of the data as a training set, and the remaining 20% as a test set. Within each fold, important features were identified using a Wilcoxon rank-sum test from the training set. From this, 44 statistically important features were used to predict the result of the test set. Four different classifiers were used for prediction: Sparse Logistic Regression, Random Forest, Gaussian Process Classifier and Support Vector Machine. Furthermore, the data was split into a training set of 100 samples and a test set of 62 samples. The classifiers were trained on the 100 samples and then applied to the 62 unknown samples to get a better measure of accuracy. Of note, our previous studies have determined that age, diet, smoking and sex are not confounders and no correction to the analyses was required [2-4]. Equally given the significant discrimination between those with PDAC and healthy controls, the data processing tools are designed such that it would disregard VOC signals that may be affected by timing of sample collection.