

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/142020>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

© 2020 Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>.



Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Volatile organic compounds (VOCs) for the non-invasive detection of pancreatic cancer from urine

Emma Daulton^a, Alfian N. Wicaksono^a, Akira Tiele^a, Hemant M Kocher^b, Silvana Debernardi^c, Tatjana Crnogorac-Jurcevic^{c,*}, James A. Covington^{a,*}

^a School of Engineering, University of Warwick, Coventry, United Kingdom

^b Centre for Tumour Biology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom

^c Centre for Cancer Biomarkers and Biotherapeutics, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom

Corresponding Author:

James A. Covington

J.A.covington@warwick.ac.uk

*Shared last authorship

Keywords (*maximum 6*):

Volatile organic compounds (VOCs), pancreatic ductal adenocarcinoma (PDAC), chronic pancreatitis (CP), urinary headspace, GC-IMS, GC-TOF-MS.

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is a particularly challenging cancer, with very low 5-year survival rates. This low survival rate is linked to late stage diagnosis, associated with the lack of approved biomarkers. One approach that is receiving considerable attention is the use of volatile organic compounds (VOCs) that emanate from biological waste as biomarkers for disease. In this study, we used urine as our biological matrix and two VOC analysis platforms: gas chromatography – ion mobility spectrometry (GC-IMS) and GC time-of-flight mass spectrometry (GC-TOF-MS). We measured the urinary headspace of samples from patients with PDAC, chronic pancreatitis (CP) and healthy controls. In total, 123 samples were tested from these groups. Results indicate that both GC-IMS and GC-TOF-MS were able to discriminate PDAC from healthy controls with high confidence and an AUC (area under the curve) in excess of 0.85. However, both methods struggled to separate CP from PDAC, with the best result of AUC 0.58. This indicates that both conditions produce similar biomarkers in the urinary headspace. Chemical identification suggests that 2,6-dimethyl-octane, nonanal, 4-ethyl-1,2-dimethyl-benzene and 2-pentanone play an important role in separating between groups. Therefore, both techniques validate this approach in identifying subjects for further investigation in a clinical setting.

1. Introduction

Cancer is currently the second leading cause of death globally, accounting for around 9.6 million deaths (1 in 6) in 2018. This trend is likely to continue, with the latest projections suggesting that annual cancer-related deaths will increase to 13 million by 2030 [1]. Thus the burden of cancer will continue to exert tremendous physical, emotional and financial strain on individuals, families, communities and healthcare systems. Pancreatic cancer is the eighth leading cause of cancer death worldwide (around 400,000 deaths) [1]. The most frequent pancreatic malignant tumour is pancreatic ductal adenocarcinoma (PDAC). PDAC represents around 85% of all reported pancreatic cancer cases [2]. It stubbornly resists our attempts to successfully target with current therapies, which are reflected in patient 5-year survival rates at below 8% [3]. This exceptionally poor prognosis is largely due to late diagnosis. However, if the disease is detected when the cancer is at an early stage (< 2 cm diameter) and still confined to the pancreas, the survival rate can increase, potentially up to 75% [4,5]. Current pancreatic cancer diagnostic techniques rely on imaging (e.g. CT and MRI), endoscopy ultrasonography and biopsy for grading tumour histology [6]. Unfortunately, these techniques are not very effective at detecting tumours that are smaller than 2 cm in diameter [7].

An alternative cancer diagnosis approach, currently attaining traction, is the use of volatile organic compounds (VOCs) that emanate from biological waste, be it breath, stool, urine or sweat [8,9]. The initial interest in the detection of VOCs for cancer has been driven by early reports on the ability of canines to identify cancer in their owners [10]. Since then, a number of studies have demonstrated that VOC profiles reflect metabolic changes in response to inflammation, necrosis, cancer development and microbiota alterations [11]. This concept has been taken forward through the use of a variety of analytical platforms, from high-end gas chromatography – mass spectrometry (GC-MS) to lower-cost/portable technologies that mimic the biological olfactory system (electronic noses) (reviewed in [12,13]). To this end, VOCs have been reported, by our and other groups, to be promising biomarkers for several cancers [14–17].

One of the simplest biological material to access is urine. This is a far less complex matrix, compared to blood, and can be collected in large volumes, completely non-invasively and with high patient acceptability (far higher than faeces). Our group have hypothesised that biomarkers for PDAC exist in urine and that they have the potential to be used as an early

detection and screening tool for patients at risk of this malignancy. To test this hypothesis, we previously investigated the urinary VOCs using a Lonestar FAIMS (field asymmetric ion mobility spectrometry) (Owlstone Medical, Cambridge, UK), which showed good discriminatory performance between PDAC and healthy control samples [18]. However, the FAIMS is unable to identify any specific biomarkers and the Lonestar unit is operated manually, testing one sample at a time, making its use more challenging in a clinical setting. Here we report on our attempts to further evaluate VOCs for the detection of PDAC, with a focus on identifying specific biomarkers of the disease and to evaluate a second gas analytical platform that is more appropriate for use in a clinical setting.

2. Materials and Methods

2.1 Urine Samples

Urine specimens utilised in this study were obtained from Barts Pancreas Tissue Bank, after patient consent and with ethical approval (reference number 13/SC/0593). The midstream urine samples were collected in 50 mL sterile containers and were kept at +4°C before freezing within four hours. In large majority of cases the containers were filled with urine with no or little headspace. All samples were stored at -80°C according to standard operating procedures compliant with tissue bank requirements under Human Tissue Act 2004. This study included patients with PDAC, chronic pancreatitis (CP) and healthy controls. Chronic pancreatitis is an inflammation of the pancreas and is a condition that is associated with an increased risk of developing pancreatic cancer [19]. This diagnostic group was included in the study to evaluate whether urinary analysis could be used to distinguish such patients at risk from the healthy and PDAC groups. In total, 123 urine samples were analysed: 33 healthy, 45 CP and 45 PDAC. The basic demographic information for the subject cohorts is summarised in Table 1.

2.2 Measurement Instrumentation

The urine samples were analysed using two analytical methods: gas chromatography – ion mobility spectrometry (GC-IMS) and by gas chromatography time-of-flight mass spectrometry (GC-TOF-MS). The GC-IMS system used was a G.A.S. GC-IMS (Dortmund, Germany) and was fitted with a 30 m, 0.32 mm inner diameter (ID) SE-54 column (CS Chromatographie Service, Langerwehe, Germany). The GC-IMS is formed of a GC pre-separator, followed by a

drift tube IMS as the detector. In use, once a sample is injected into the device, the GC selectively slows down molecules based on the interaction between the molecule and the GC columns retentive coating, as they pass through the column, and thus temporally separates them as they exit. These enter a drift-tube, where the molecules are ionised (in this case, using a low-radiation tritium source). The ionised sample is then passed through a drift tube, to which a known electric field is applied to propel the ions along it. Against the flow of ions, a buffer gas is passed that collides with the ions. Large ions have many collisions and slow down, whilst small ions have less collisions and maintain their momentum. Therefore, the time taken for ions to be detected is based on their interaction with the electric field and the buffer gas [20].

GC-TOF-MS works in a similar way to traditional GC-MS methods, but instead of filtering ions by mass, the TOF utilises ‘time of flight’ and analyses all ions present. The GC-TOF-MS system consists of a TRACE 1300 GC (Thermo Fisher Scientific, Loughborough, UK), combined with a BenchTOF-HD TOF-MS (Markes Intl., Llantrisant, UK). The GC column used was a 20 m, 0.18 mm ID, Rxi-624Sil MS column (Thames Restek, Saunderton, UK). This system also includes a high-throughput autosampler and a thermal desorption unit, ULTRA-xr and UNITY-xr, respectively (both from Markes Intl.). As with a GC-MS system, samples are injected into the GC column. Compounds within a sample can be separated due to the interaction of the molecules with the stationary phase of the column (as with GC-IMS). The separated molecules leave the column at different times, giving a retention time. The molecules are then ionised, and the resultant ions are detected by the mass spectrometer. In a GC-TOF-MS system, once the sample has passed through the GC column, the separated molecules are ionised and enter the TOF ‘flight box’. TOF-MS works on the principle that when molecules are exposed to a pre-determined amount of energy, larger, heavier ions will take longer than smaller, lighter ions to travel a set distance. The addition of the TOF into the GC-MS method allows for further separation of the ions present in a sample, increasing the sensitivity of the method [21].

2.3 GC-IMS Measurements

For GC-IMS measurements, we have developed a standard method for testing urine samples, where we have optimised the sample temperature, method temperatures and flow rates to maximise information content and produce reproducible results. The urine samples were shipped on dry ice to the University of Warwick and were stored at -20°C. Prior to analysis,

the samples were thawed for four hours at room temperature. 5 ml samples of urine were aliquoted into 20 ml glass vials and sealed with a crimp cap and septa. Once sealed, the samples were agitated and heated to 40°C for 10 min. For sampling, a sterile needle, attached to the heated sample inlet of the GC-IMS, was inserted into the sample headspace, through the septa. The GC-IMS sampled 2 ml of the headspace for analysis. The sampling and analysis were performed using the following settings: GC flow rate = 20 ml/min, drift tube flow rate = 150 ml/min, IMS temperature = 45°C, GC temperature (fixed) = 45°C, Sample loop = 45°C and inlet injector = 45°C. The analysis time for each sample was 8 min. For quality control, all flow rates, method temperatures and RIP magnitude and location were checked for each sample, to be within unit tolerances. Furthermore, the output obtained from each sample was visually checked to ensure that they contained the expected level of information content. Finally, air blanks were run either side of a test batch of samples (20 samples) to ensure that there was no machine drift

2.4 GC-TOF-MS Measurements

For GC-TOF-MS analysis, urine samples were defrosted as described in section 2.3, with 10 ml of sample aliquoted into a 20 ml glass vial, which were then sealed with a specially adapted septa and crimp cap. A thermal desorption (TD) sorbent tube (C2-AXXX-5149, Markes Intl., Llantrisant, UK) was placed through the septa and into the headspace above the urine sample. The vial and sorbent tube were then placed into a heater block and heated to 40°C for 1 hour. Once completed, the tube was removed from the top of the vial and placed into the autosampler for analysis. The ULTRA-xr was set to run with a stand-by split of 150°C with an overlap (this allows the auto-sampler to reduce the overall run time), a GC run-time of 30 min, and a minimum carrier pressure of 5 psi. For each sample there is a pre-purge of 1 min. The tube was desorbed for 10 min at 250°C, with the trap purge time set to 1 min, and the trap cooled to -30°C. The trap was then purged for 3 min at a temperature of 300°C. The GC-TOF-MS method measured masses from 45 to 500 amu (atomic mass units). The transfer line and ion source are both heated to 250°C, with an ionisation voltage of -70.000V. The GC oven was heated to 280°C for 25 min. Upon completion of the GC-TOF-MS run-time, peaks were identified using NIST list 2016. For quality control, the machine was calibrated in line with the manufacturer's recommendation. The method used was developed using healthy control urine samples from a previous study, where parameters were optimised to maximise separation and chemical content.

2.5 Data Analysis

GC-IMS generate very high-dimensional data (typically 11 million data points per sample). Therefore, to aid in data analysis, pre-processing steps were applied to reduce its dimensionality. As the chemical information on the output is located around the central region of the output data, we can crop this section and discard the rest. The size of the crop area is chosen through visual observation of the data and the same crop values are used for all the data. The second step is to delete the reactant ion peak (RIP), which is output when there are no chemicals present. This is undertaken by selecting a line where there is no chemical information (though contains the RIP) and subtracting this from the rest of the data. A small threshold is added to this value to remove the background noise, which was selected manually as applied the same to all the data. These steps reduce the number of data points from around 11 million to 10,000, without losing any volatile information.

The data was then analysed using a 10-fold cross validation, using our custom developed pipeline, created in 'R' (version 3.6.1). In this case, the data is divided into two groups, with 90% of the data being used as a training set and 10% is used as a test set. To identify features (data points) that hold discriminatory information, a rank-sum test was undertaken and features with the lowest p-values are selected from training. These features are used for training three different classifiers, specifically random forest (RF), Gaussian process (GP), and sparse logistic regression (SLR). We have previously used these successfully in a range of medical VOC studies [22–24]. Using the developed model, the test samples are then classified and a probability for each test sample is created. This process was repeated 10 times so that all the samples were classified as test samples. Using the resultant probabilities, statistical parameters, including area under the curve (AUC), sensitivity, specificity, negative predictive value and positive predictive value are calculated.

The GC-TOF-MS analysis was undertaken using a similar approach. In this case, each identified chemical was used as a feature and the abundance as the magnitude of the feature. This was processed in the same way as described above, but without the rank-sum test as the data already had low dimensionality. From this process, we are also able to identify which features/chemicals hold discriminatory information.

3. Results

A typical output plot from the GC-IMS and GC-TOF-MS are shown in Figure 1 and Figure 2, respectively. For the GC-IMS, the background is blue, with the RIP line being red (which is always present). The chemicals are seen as white/red areas. The colour represents the intensity of the chemical. Three different analyses were undertaken comparing PDAC, healthy and CP groups. GC-IMS analysis indicated that there were differences between the three groups. The three classifiers resulted in very similar data, with the best results shown in Table 2 and the complete data in supplementary table TS1. The PDAC ROC curves are shown in Figure 3. The results indicate that GC-IMS (sensitivity 84%, specificity 94%, p -value >0.0001) can separate PDAC from healthy controls. However, the separation between PDAC and CP group is not as pronounced (sensitivity 51%, specificity 73%, p -value 0.11), indicating that the same biomarkers are involved in both conditions.

We undertook the same analytical methods for the data analysed by GC-TOF-MS. In this case, the statistical results are shown in Table 3, with the PDAC ROC curves in Figure 4. Here, we were also able to separate PDAC from healthy controls (sensitivity 72%, specificity 96%, p -value >0.0001). However, when comparing CP to PDAC (sensitivity 38%, specificity 88%, p -value 0.28), our results did not show a significant difference.

From the analysis, we were able to identify chemicals that held discriminatory properties. 15 chemicals for each comparison were identified. From this list, a rank-sum test of each chemical/features was undertaken and the top three chemicals for each comparison are listed in Table 4, with the complete set in supplementary information TS3. As shown, we have identified common chemicals from each of the analyses, with 2,6-dimethyl-octane, 2-pentanone, nonanal and 4-ethyl-1,2-dimethyl-benzene being most frequent.

4. Discussion

In a recent review paper, Bax et al. [11] evaluated and compared cancer biomarker trends in urine as a new diagnostic pathway. Five studies [25–29] investigating urinary pancreatic cancer biomarkers were included in the review. The employed technologies were nuclear magnetic resonance (NMR) spectroscopy and gel electrophoresis liquid chromatography-tandem mass spectrometry (GeLC-MS/MS). To the best of our knowledge, this was the first urinary study

to utilise GC-IMS and GC-TOF-MS technologies to investigate PDAC, CP and healthy controls. Both the GC-IMS and GC-TOF-MS were able to separate PDAC from healthy controls with good sensitivity and specificity, with the GC-IMS outperforming the GC-TOF-MS. Only the GC-IMS was able to accurately separate CP from controls, whilst the GC-TOF-MS showed inferior results. Comparing PDAC with CP for both methods, the AUCs were around 0.6, showing a modest diagnostic performance. This limited ability to separate these two groups, could well have impacted the overall diagnostic performance of PDAC vs all other samples, with both methods having an AUC of around 0.7. In statistical results, these variations are likely to be associated with the choice of column between the two analytical platforms and the sample capture process. The sorbent tubes used for the GC-TOF-MS analysis capture analytes from C₃ (though more likely C₄) and upwards, whilst the GC-IMS is analysing all the chemicals through direct headspace injection. This is a limitation of the experimental method, where the absorbent tube and trap on the TD unit cannot go below C₃ (with C referring to the number of carbon atoms). Another difference between the platforms is that the GC-IMS molecular detection depends on the proton affinity of the molecule. However, the overall statistical performance was similar. Some of the discriminatory VOCs, identified in this study, have been suggested as potential biomarkers in breath or stool for other disease. For example, 2-pentanone has been found to be associated with several diseases, such as non-alcoholic fatty liver disease [30], inflammatory bowel disease (ulcerative colitis and Crohn's disease) [31,32] and lung cancer [33]. Nonanal is a saturated fatty aldehyde formally arising from reduction of the carboxy group of nonanoic acid. It has been observed as a discriminatory VOC for other cancers, such as ovarian and lung cancer [34,35]. This indicates that some of these compounds might be more generic markers of inflammation or illness. As stated above, a limitation of this study is that the analytical approaches we have used here may have missed other potential biomarkers. For example, biomarkers under C₃ could be critical. Only undertaking additional experiments with direct headspace injection into a GC-MS or similar **technique** will allow us to identify if such biomarkers exist. Though GC-IMS will measure these chemicals, identification is more limited, and it will not detect those molecules with low proton affinity. In addition, due to difficulty in obtaining these samples, we were not able to undertake extensive optimisation of the sampling conditions for this specific disease and we relied on a method we had previously developed for the analysis of urine samples in general. Therefore, better diagnostic performance could be achieved in the future if more urine samples could be collected. Furthermore, the total number of samples was only 123. In the follow up study, we

will be analysing more samples, which will enable us to divide the samples into a separate test and training set. This will give us a fully verified result.

5. Conclusion

In this paper we investigated the use of urinary headspace volatiles to identify patients suffering from PDAC. Urinary headspace was analysed by GC-IMS and GC-TOF-MS, with both showing potential in separating PDAC from healthy controls. However, both GC-IMS and GC-TOF-MS showed only small differences between PDAC and CP, indicating that there is commonality between the VOCs produced by both conditions. This is supported by the high sensitivity of CP vs healthy controls. Chemical identification suggests that 2,6-dimethyl-octane, nonanal, 4-ethyl-1,2-dimethyl-benzene and 2-pentanone play an important role in separating the data into disease groups or controls. Further work is needed to validate these biomarkers in a larger study. However, we believe this approach might hold a potential as a completely non-invasive detection tool for pancreatic cancer.

Acknowledgements

This research received support from CRUK Barts Centre Pump-priming funds. The PhD of Alfian N. Wicaksono is supported by Indonesian Endowment Fund for Education (LPDP), Ministry of Finance, Republic of Indonesia

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA. Cancer J. Clin.* 68 (2018) 394–424. <https://doi.org/10.3322/caac.21492>.
- [2] I. Aier, R. Semwal, A. Sharma, P.K. Varadwaj, A systematic assessment of statistics, risk factors, and underlying features involved in pancreatic cancer, *Cancer Epidemiol.* 58 (2019) 104–110. <https://doi.org/10.1016/j.canep.2018.12.001>.
- [3] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2016, *CA. Cancer J. Clin.* 66 (2016) 7–30. <https://doi.org/10.3322/caac.21332>.
- [4] S.T. Chari, K. Kelly, M.A. Hollingsworth, S.P. Thayer, D.A. Ahlquist, D.K. Andersen, S.K. Batra, T.A. Brentnall, M. Canto, D.F. Cleeter, M.A. Firpo, S.S. Gambhir, V.L.W. Go, O.J. Hines, B.J. Kenner, D.S. Klimstra, M.M. Lerch, M.J. Levy, A. Maitra, S.J. Mulvihill, G.M. Petersen, A.D. Rhim, D.M. Simeone, S. Srivastava, M. Tanaka, A.I. Vinik, D. Wong, **Early detection of sporadic pancreatic cancer**, *Pancreas*. 44 (2015) 693–712. <https://doi.org/10.1097/MPA.0000000000000368>.
- [5] A. Kanno, A. Masamune, K. Hanada, H. Maguchi, Y. Shimizu, T. Ueki, O. Hasebe, T. Ohtsuka, M. Nakamura, M. Takenaka, M. Kitano, M. Kikuyama, T. Gabata, K. Yoshida, T. Sasaki, M. Serikawa, T. Furukawa, A. Yanagisawa, T. Shimosegawa, Multicenter study of early pancreatic cancer in Japan, *Pancreatology*. 18 (2018) 61–67. <https://doi.org/10.1016/j.pan.2017.11.007>.
- [6] A.S. Takhar, P. Palaniappan, R. Dhingsa, D.N. Lobo, Recent developments in diagnosis of pancreatic cancer, *BMJ*. 329 (2004) 668–673. <https://doi.org/10.1136/bmj.329.7467.668>.
- [7] G. Xie, L. Lu, Y. Qiu, Q. Ni, W. Zhang, Y.-T. Gao, H.A. Risch, H. Yu, W. Jia, Plasma **Metabolite biomarkers for the detection of pancreatic cancer**, *J. Proteome Res.* 14 (2015) 1195–1202. <https://doi.org/10.1021/pr501135f>.
- [8] S. Sethi, R. Nanda, T. Chakraborty, **Clinical application of volatile organic compound analysis for detecting infectious diseases**, *Clin. Microbiol. Rev.* 26 (2013) 462–475. <https://doi.org/10.1128/CMR.00020-13>.
- [9] J. Dummer, M. Storer, M. Swanney, M. McEwan, A. Scott-Thomas, S. Bhandari, S. Chambers, R. Dweik, M. Epton, Analysis of biogenic volatile organic compounds in human health and disease, *TrAC Trends Anal. Chem.* 30 (2011) 960–967.

<https://doi.org/10.1016/j.trac.2011.03.011>.

- [10] M. McCulloch, T. Jezierski, M. Broffman, A. Hubbard, K. Turner, T. Janecki, **Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers**, *Integr. Cancer Ther.* 5 (2006) 30–39. <https://doi.org/10.1177/1534735405285096>.
- [11] Bax, Lotesoriere, Sironi, Capelli, **Review and comparison of cancer biomarker trends in urine as a basis for new diagnostic pathways**, *Cancers (Basel)*. 11 (2019) 1244. <https://doi.org/10.3390/cancers11091244>.
- [12] R.P. Arasaradnam, J.A. Covington, C. Harmston, C.U. Nwokolo, Review article: next generation diagnostic modalities in gastroenterology – gas phase volatile compound biomarker detection, *Aliment. Pharmacol. Ther.* 39 (2014) 780–789. <https://doi.org/10.1111/apt.12657>.
- [13] W. Wojnowski, T. Dymerski, J. Gębicki, J. Namieśnik, **Electronic Noses in medical diagnostics**, *Curr. Med. Chem.* 26 (2019) 197–215. <https://doi.org/10.2174/0929867324666171004164636>.
- [14] M.M. Widlak, M. Neal, E. Daulton, C.L. Thomas, C. Tomkins, B. Singh, C. Harmston, A. Wicaksono, C. Evans, S. Smith, R.S. Savage, J.A. Covington, R.P. Arasaradnam, Risk stratification of symptomatic patients suspected of colorectal cancer using faecal and urinary markers, *Color. Dis.* 20 (2018) O335–O342. <https://doi.org/10.1111/codi.14431>.
- [15] E. Westenbrink, R.P. Arasaradnam, N. O’Connell, C. Bailey, C. Nwokolo, K.D. Bardhan, J.A. Covington, Development and application of a new electronic nose instrument for the detection of colorectal cancer, *Biosens. Bioelectron.* 67 (2015) 733–738. <https://doi.org/10.1016/j.bios.2014.10.044>.
- [16] M. Cauchi, C.M. Weber, B.J. Bolt, P.B. Spratt, C. Bessant, D.C. Turner, C.M. Willis, L.E. Britton, C. Turner, G. Morgan, Evaluation of gas chromatography mass spectrometry and pattern recognition for the identification of bladder cancer from urine headspace, *Anal. Methods*. 8 (2016) 4037–4046. <https://doi.org/10.1039/C6AY00400H>.
- [17] T. Khalid, R. Aggio, P. White, B. De Lacy Costello, R. Persad, H. Al-Kateb, P. Jones, C.S. Probert, N. Ratcliffe, **Urinary volatile organic compounds for the detection of prostate cancer**, *PLoS One*. 10 (2015) e0143283. <https://doi.org/10.1371/journal.pone.0143283>.
- [18] R.P. Arasaradnam, A. Wicaksono, H. O’Brien, H.M. Kocher, J.A. Covington, T. Crnogorac-Jurcevic, **Noninvasive diagnosis of pancreatic cancer through detection of**

- volatile organic compounds in urine, *Gastroenterology*. 154 (2018) 485–487. <https://doi.org/10.1053/j.gastro.2017.09.054>.
- [19] J. Kirkegård, F.V. Mortensen, D. Cronin-Fenton, **Chronic pancreatitis and pancreatic cancer risk: A systematic review and meta-analysis**, *Am. J. Gastroenterol.* 112 (2017) 1366–1372. <https://doi.org/10.1038/ajg.2017.218>.
- [20] J.I. Baumbach, Process analysis using ion mobility spectrometry, *Anal. Bioanal. Chem.* 384 (2006) 1059–1070. <https://doi.org/10.1007/s00216-005-3397-8>.
- [21] B. Aebi, R. Sturny-Jungo, W. Bernhard, R. Blanke, R. Hirsch, Quantitation using GC–TOF-MS: example of bromazepam, *Forensic Sci. Int.* 128 (2002) 84–89. [https://doi.org/10.1016/S0379-0738\(02\)00165-2](https://doi.org/10.1016/S0379-0738(02)00165-2).
- [22] L. Lacey, E. Daulton, A. Wicaksono, J.A. Covington, S. Quenby, Detection of Group B *Streptococcus* in pregnancy by vaginal volatile organic compound analysis: a prospective exploratory study, *Transl. Res.* 216 (2020) 23–29. <https://doi.org/10.1016/j.trsl.2019.09.002>.
- [23] A. Tiele, A. Wicaksono, E. Daulton, E. Ifeakor, V. Eyre, S. Clarke, L. Timings, S. Pearson, J. Covington, X. Li, Breath-based non-invasive diagnosis of Alzheimer’s disease: A pilot study, *J. Breath Res.* (2019). <https://doi.org/10.1088/1752-7163/ab6016>.
- [24] M.D. Rouvroye, A. Wicaksono, S. Bosch, E. Savelkoul, J.A. Covington, H. Beaumont, C.J. Mulder, G. Bouma, T.G.J. de Meij, N.K.H. de Boer, **Faecal scent as a novel non-invasive biomarker to discriminate between coeliac disease and refractory coeliac disease: A proof of principle study**, *Biosensors*. 9 (2019) 69. <https://doi.org/10.3390/bios9020069>.
- [25] C. Napoli, N. Sperandio, R.T. Lawlor, A. Scarpa, H. Molinari, M. Assfalg, **Urine metabolic signature of pancreatic ductal adenocarcinoma by ¹H Nuclear Magnetic Resonance: identification, mapping, and evolution**, *J. Proteome Res.* 11 (2012) 1274–1283. <https://doi.org/10.1021/pr200960u>.
- [26] V.W. Davis, D.E. Schiller, D. Eurich, O.F. Bathe, M.B. Sawyer, **Pancreatic ductal adenocarcinoma is associated with a distinct urinary metabolomic signature**, *Ann. Surg. Oncol.* 20 (2013) 415–423. <https://doi.org/10.1245/s10434-012-2686-7>.
- [27] E.R. Luszczek, J.A. Paulo, J.R. Saltzman, V. Kadiyala, P.A. Banks, G. Beilman, D.L. Conwell, **Urinary ¹H-NMR metabolomics can distinguish pancreatitis patients from healthy controls**, *J. Pancreas.* 14 (2013) 161–170. <https://doi.org/10.6092/1590-8577/1294>.
- [28] T.P. Radon, N.J. Massat, R. Jones, W. Alrawashdeh, L. Dumartin, D. Ennis, S.W. Duffy,

- H.M. Kocher, S.P. Pereira, L. Guarner (posthumous), C. Murta-Nascimento, F.X. Real, N. Malats, J. Neoptolemos, E. Costello, W. Greenhalf, N.R. Lemoine, T. Crnogorac-Jurcevic, **Identification of a three-biomarker panel in urine for early detection of pancreatic adenocarcinoma**, *Clin. Cancer Res.* 21 (2015) 3512–3521. <https://doi.org/10.1158/1078-0432.CCR-14-2467>.
- [29] J. Mayerle, H. Kalthoff, R. Reszka, B. Kamlage, E. Peter, B. Schniewind, S. González Maldonado, C. Pilarsky, C.-D. Heidecke, P. Schatz, M. Distler, J.A. Scheiber, U.M. Mahajan, F.U. Weiss, R. Grützmann, M.M. Lerch, Metabolic biomarker signature to differentiate pancreatic ductal adenocarcinoma from chronic pancreatitis, *Gut*. 67 (2018) 128–137. <https://doi.org/10.1136/gutjnl-2016-312432>.
- [30] S. Van den Velde, F. Nevens, P. Van hee, D. van Steenberghe, M. Quirynen, GC–MS analysis of breath odor compounds in liver patients, *J. Chromatogr. B.* 875 (2008) 344–348. <https://doi.org/10.1016/j.jchromb.2008.08.031>.
- [31] C.E. Garner, S. Smith, B. de Lacy Costello, P. White, R. Spencer, C.S.J. Probert, N.M. Ratcliffe, Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease, *FASEB J.* 21 (2007) 1675–1688. <https://doi.org/10.1096/fj.06-6927com>.
- [32] I. Ahmed, R. Greenwood, B. Costello, N. Ratcliffe, C.S. Probert, Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease, *Aliment. Pharmacol. Ther.* 43 (2016) 596–611. <https://doi.org/10.1111/apt.13522>.
- [33] Y. Saalberg, M. Wolff, VOC breath biomarkers in lung cancer, *Clin. Chim. Acta.* 459 (2016) 5–9. <https://doi.org/10.1016/j.cca.2016.05.013>.
- [34] H. Amal, D.-Y. Shi, R. Ionescu, W. Zhang, Q.-L. Hua, Y.-Y. Pan, L. Tao, H. Liu, H. Haick, Assessment of ovarian cancer conditions from exhaled breath, *Int. J. Cancer.* 136 (2015) E614–E622. <https://doi.org/10.1002/ijc.29166>.
- [35] 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. <https://doi.org/10.1002/cam4.162>.

Table 1: Demographics of sample groups and cancer stage

Group	Gender (Male:Female)	Av. age (range) years	Number of subjects (Stage S x Number)
PDAC	22M:23F	64.1 +/- 10.7 (29-77)	45 (SI x 1; SII x 21; SIII x 20; SIV x 3)
CP	28M:17F	52.9 +/- 12.2 (21-78)	45 (not applicable)
Healthy	15:M:18F	49.9 +/- 7.8 (30-66)	33 (not applicable)

Table 2: Statistical analysis of GC-IMS data (best results)

Statistical parameter	PDAC vs All	PDAC vs CP	PDAC vs Healthy	CP vs Healthy
Best classifier	SLR	GP	SLR	GP
AUC	0.69 (0.58 – 0.79)	0.58 (0.45 – 0.71)	0.88 (0.79 – 0.97)	0.86 (0.77 – 0.95)
Sensitivity	0.72 (0.56 – 0.85)	0.51 (0.35 – 0.67)	0.84 (0.69 – 0.73)	0.80 (0.64 – 0.91)
Specificity	0.60 (0.57 – 0.71)	0.73 (0.56 – 0.85)	0.94 (0.79 – 0.99)	0.91 (0.75 – 0.98)
PPV	0.52	0.67	0.95	0.91
NPV	0.78	0.58	0.81	0.78
p-value	4.39×10^{-4}	0.11	1.18×10^{-8}	9.79×10^{-9}

Table 3: Statistical analysis of GC-TOF-MS data (best results)

Statistical parameter	PDAC vs All	PDAC vs CP	PDAC vs Healthy	CP vs Healthy
Best classifier	SLR	RF	SLR	RF
AUC	0.75 (0.63 – 0.87)	0.55 (0.37 – 0.73)	0.86 (0.75 – 0.97)	0.67 (0.50 – 0.83)
Sensitivity	0.52 (0.37 – 0.67)	0.38 (0.18 – 0.62)	0.72 (0.51 – 0.88)	0.38 (0.18 – 0.62)
Specificity	0.96 (0.79 – 1)	0.88 (0.68 – 0.97)	0.96 (0.79 – 1)	0.96 (0.80 – 1.00)
PPV	0.96	0.73	0.95	0.89
NPV	0.51	0.62	0.77	0.65
p-value	3.11×10^{-4}	0.28	1.81×10^{-6}	2.75×10^{-2}

Table 4: Chemicals used to separate sample groups.

PDAC vs CP	PDAC vs Healthy	CP vs Healthy
2-pentanone	2,6-dimethyl-octane	2-pentanone
Nonanal	Nonanal	Benzene, 1-ethenyl-2-methyl-
4-ethyl-1,2-dimethyl-Benzene	4-ethyl-1,2-dimethyl-Benzene	4-ethyl-1,2-dimethyl-Benzene

List of Figures:

Figure 1: GC-IMS output of a PDAC urine sample (*in colour*)

Figure 2: Typical GC-TOF-MS output of a PDAC urine sample, with example chemicals labelled.

Figure 3: ROC for (a) PDAC vs Healthy and (b) PDAC vs CP using GC-IMS

Figure 4: ROC for (a) PDAC vs Healthy and (b) PDAC vs CP using GC-TOF-MS