

# SCIENTIFIC REPORTS

OPEN

## Discrimination between the human prostate normal and cancer cell exometabolome by GC-MS

Ana Rita Lima<sup>1</sup>, Ana Margarida Araújo<sup>1</sup>, Joana Pinto<sup>1</sup>, Carmen Jerónimo<sup>2,3</sup>, Rui Henrique<sup>2,3,4</sup>, Maria de Lourdes Bastos<sup>1</sup>, Márcia Carvalho<sup>1,5</sup> & Paula Guedes de Pinho<sup>1</sup>

Serum prostate-specific antigen (PSA) is currently the most used biomarker in clinical practice for prostate cancer (PCa) detection. However, this biomarker has several drawbacks. In this work, an untargeted gas chromatography-mass spectrometry (GC-MS)-based metabolomic profiling of PCa cells was performed to prove the concept that metabolic alterations might differentiate PCa cell lines from normal prostate cell line. For that, we assessed the differences in volatile organic compounds (VOCs) profile in the extracellular medium (exometabolome) of four PCa cell lines and one normal prostate cell line at two pH values (pH 2 and 7) by GC-MS. Multivariate analysis revealed a panel of volatile metabolites that discriminated cancerous from normal prostate cells. The most altered metabolites included ketones, aldehydes and organic acids. Among these, we highlight pentadecane-2-one and decanoic acid, which were significantly increased in PCa compared to normal cells, and cyclohexanone, 4-methylheptan-2-one, 2-methylpentane-1,3-diol, 4-methylbenzaldehyde, 1-(3,5-dimethylfuran-2-yl) ethanone, methyl benzoate and nonanoic acid, which were significantly decreased in PCa cells. The PCa volatilome was markedly influenced by the VOCs extraction pH, though the discriminant capability was similar. Overall, our data suggest that VOCs monitoring has the potential to be used as a PCa screening methodology.

Prostate cancer (PCa) is the second leading cause of cancer-related death in men in most western countries<sup>1</sup> and for 2017, 26,730 resulting deaths are predicted for USA alone<sup>2</sup>. Prostate-specific antigen (PSA) is currently the most used biomarker for PCa detection, in combination with digital rectal examination (DRE)<sup>3</sup>. However, the role of serum PSA levels as a screening tool for PCa meets with important limitations. Although raised PSA levels (i.e., >4.0 ng/mL) are undoubtedly associated with the presence of PCa, benign conditions, such as prostatitis and benign prostatic hyperplasia (BPH), which are common in the elderly, also cause elevated serum PSA. Consequently, there is a relatively high frequency of unnecessary prostate biopsies, an invasive procedure which is expensive and uncomfortable for the cancer suspects<sup>4</sup>. Moreover, a significant proportion of men with PSA levels within the normal range harbor PCa (up to 22%) and a significant number of these show pathologic features of tumor aggressiveness<sup>5-8</sup>. Indeed, 21% of men with positive end of study biopsies enrolled in the Prostate Cancer Prevention Trial (PCPT) had serum PSA levels between 2.6 and 3.9 ng/mL and 15.4% of the tumors found in men with PSA levels < 2.5 ng/mL were high grade cancers<sup>9</sup>. The PSA test for PCa diagnosis has an area under the curve (AUC) of 0.682 and considering a cut-off of 4.1 ng/ml this biomarker shows a specificity of 93.8% and a sensitivity of 20.5%<sup>1</sup>. Furthermore, PSA is unable to differentiate aggressive from indolent PCa, which may lead to overtreatment<sup>10,11</sup>. Indeed, a large US-based trial found no benefit of PSA screening in reducing PCa-related mortality<sup>12</sup> and even the large European trial that found a moderate benefit (approx. 20%) acknowledged that 1410 men would have to be screened and additional 48 cases of PCa would need to be treated to avoid a single death from PCa<sup>13</sup>. Besides the human cost, implementing widespread PSA screening for PCa might also lead to

<sup>1</sup>UCIBIO/REQUIMTE, Department of Biological Sciences, Laboratory of Toxicology, Faculty of Pharmacy, University of Porto, Porto, Portugal. <sup>2</sup>Cancer Biology & Epigenetics Group, Research Center (CI-IPOP) Portuguese Oncology Institute of Porto (IPO Porto), Porto, Portugal. <sup>3</sup>Department of Pathology and Molecular Immunology-Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal. <sup>4</sup>Department of Pathology, Portuguese Oncology Institute of Porto (IPO Porto), Porto, Portugal. <sup>5</sup>UFP Energy, Environment and Health Research Unit (FP-ENAS), University Fernando Pessoa, Porto, Portugal. Correspondence and requests for materials should be addressed to A.R.L. (email: [ritaclima@hotmail.com](mailto:ritaclima@hotmail.com)) or P.G. (email: [pguedes@ff.up.pt](mailto:pguedes@ff.up.pt))

double of the total financial costs associated with PCa management<sup>14</sup>. Due to these limitations, the use of PSA for populational screening has been challenged<sup>15</sup>.

Although several molecular tests have been developed over the years and some have already obtained FDA approval<sup>16</sup>, accurate early detection of PCa remains an unmet need. Thus, discovery and validation of novel, more specific and cost-effective biomarkers that might improve early PCa diagnosis and more precisely forecast its clinical behavior in an individual basis is an important research aim.

Metabolomics is a powerful analytical tool in oncology, endowing novel biomarkers and therapeutic targets, as cancer cells have the capacity to modify many homeostatic systems within the body and, consequently, change the production, use and levels of many metabolites<sup>17,18</sup>. A metabolomic approach may allow for the discovery of biochemical signatures, and, consequently, of differences between cancer and healthy metabolic phenotypes<sup>17</sup> in non-invasive samples. An early intervention is possible using metabolomics, since it is believed that metabolic alterations precede neoplastic proliferation<sup>11</sup>.

Several different matrices may be used in PCa metabolomics studies, but the most common are biofluids (e.g., urine and serum/plasma), tissues and cell lines. Generally, when the chosen matrix is a biofluid and/or a tissue it is very important to be aware that metabolic profile can be altered by factors not related to cancer cell metabolism, like age, diet, drugs, chronobiological variations, among others, which are very important to control to obtain reliable results. On the other hand, experiments performed in cultured cell lines have several advantages compared to the use of urine or plasma, overcoming these problems<sup>17</sup>. In fact, the use of cell lines in preliminary studies have important advantages, as cell lines allow to circumvent several important confounding factors, like age, diet, drugs, chronobiological variations, among others<sup>19–21</sup>. In addition, cell lines have a perfectly defined cell state which allows for the analysis of a target metabolic status. So, we believe that cell lines are the ideal matrix for hypothesis generation and to unveil unchanged metabolic signature originated directly from cells, metabolic alterations that do not appear in studies using animal models or human subjects, due to sample biological complexity, may be revealed<sup>19–21</sup>. Notwithstanding, this *in vitro* model presents some limitations, particularly the fact that cultured cells fail to reproduce the complex cell–cell and cell–matrix interactions in the tumour microenvironment and these interactions are very important for metabolic alterations occurring with tumor progression<sup>20,22</sup>. For these reasons, the findings obtained in *in vitro* studies need further confirmation and validation in biofluids from PCa patients.

Volatilome is defined as the analysis of the volatile profile of a biological system, being accomplished by the evaluation of the volatile organic compounds (VOCs)<sup>23</sup>. VOCs correspond to a carbon-based chemical group, with low molecular weight and high volatility. These compounds are excreted (exometabolome) from the human body in several tissues, accumulated and/or eliminated and can reflect the metabolic condition of an individual<sup>24,25</sup> and, therefore, have the potential to provide new biomarkers for cancer detection. Since VOCs are the end products of cell metabolism, the volatilome can provide information about intracellular metabolic status. Alterations in VOCs profile may be related to modifications in gene activation, gene expression, proteins and activity of enzymes involved in metabolic pathways<sup>26,27</sup>.

Importantly, the volatilome analysis does not require cell disruption, which allows the use of the same cells at different times<sup>28</sup>. The most commonly used method to analyse VOCs is the Headspace-Solid Phase Microextraction/Gas Chromatography-Mass Spectrometry (HS-SPME/GC-MS), which involves simple and rapid sample preparation, allowing high-throughput screening, has a good sensitivity and does not require a concentration step before analysis<sup>29,30</sup>. Moreover, VOCs can be detected in several different matrices, including exhaled breath and urine, particularly interesting for clinical practice use due to their non-invasive nature<sup>23,31</sup>.

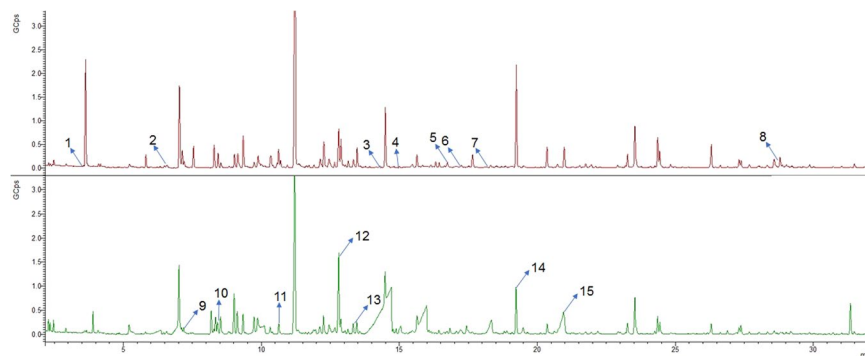
Some reports have already focused on the application of VOCs analysis in different cancers, namely breast<sup>32</sup>, lung<sup>31</sup>, head and neck<sup>33</sup>, esophago-gastric<sup>34</sup>, colorectal skin<sup>35</sup>, skin<sup>36</sup>, liver<sup>37</sup> and renal<sup>23</sup>, proving to be able to discriminate cancer from normal samples. With great interest to this study, some reports determining VOCs in urine of PCa patients showed significant differences from urines of control individuals<sup>38,39</sup>.

The main objective of this study is to perform the volatile profiling of four different PCa cell lines (22RV1, PC3; DU145; LNCaP) and one normal prostate epithelial cell line (PNT2) to prove the concept that VOCs emanated into the extracellular medium may discriminate PCa cells from normal prostate cells. VOCs extraction is performed at two different pH values, pH 7 (natural pH of prostate cells' medium) and pH 2 because it is well established that tumor microenvironment is acidic to promote tumor progression and metastasis<sup>40</sup>. This strategy allows to obtain a more comprehensive evaluation of PCa volatile profile and to infer which pH is optimum for VOCs extraction and detection. VOCs significantly altered as a consequence of PCa metabolism will integrate a panel of candidate biomarkers for early detection of PCa that may, in the future, be translated and validated in other biological matrices (e.g. urine) and used in clinical practice. Furthermore, cell lines included in this study exhibit different hormone-dependency and potential for invasiveness, which allowed for extending our investigation on the metabolites responsible for discriminating PCa cell lines according to tumour androgen-deprivation sensitivity and aggressiveness. To the best of our knowledge, this work is the first to analyze VOCs profile in human PCa cell lines.

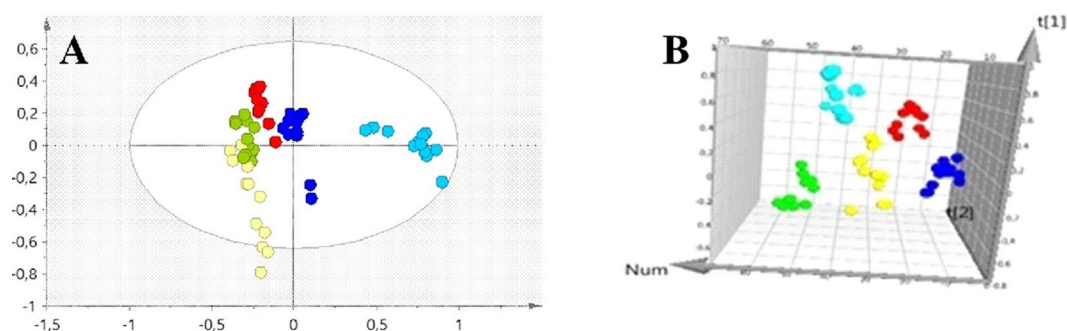
## Results

In this study, we performed a metabolomic untargeted approach, at two different pHs, pH 7 and pH 2, to evaluate alterations in the volatilome of PCa cell lines (22RV1, PC3, DU145 and LNCaP) when compared with normal prostate cell line (PNT2). A multivariate analysis (MVA) approach was applied to evaluate the ability of volatilome to discriminate the different cell lines used in this study. Figure 1 shows the representative chromatograms of the volatile profile obtained at pH 7 and pH 2. A total of 239 and of 221 features were detected in the chromatograms obtained at pH 7 and pH 2 samples, respectively (Fig. 1).

The reproducibility of the analytical method is evidenced by the QCs cluster in PCA scores plot (Supplementary Fig. 1). Furthermore, the MVA analysis (principal component analysis (PCA) and partial least



**Figure 1.** Chromatograms from quality control samples (QCs) at pH 7 and pH 2 (1: 3-methylbut-3-en-2-ol; 2: 1,4-xylene; 3: 2,7-dimethyloctan-1-ol; 4: 2-(4-methylcyclohex-3-en-1-yl)propan-2-ol; 5: 3,7-dimethyloct-7-en-1-ol; 6: 1,3-benzothiazol; 7: decan-1-ol; 8: pentadecan-2-one; 9: cyclohexanone; 10: 4-methylheptan-2-one; 11: 2-methylpentan-1,3-diol; 12: 4-methylbenzaldehyde; 13: methyl benzoate; 14: nonanoic acid; 15: decanoic acid).



**Figure 2.** PCA scores scatter plot obtained for the HS-SPME/GC-MS chromatograms of all samples (PNT2: green; 22RV1: dark blue; PC3: light blue; DU145: red; LNCaP: yellow) (A) at pH 7 ( $R^2X = 0.445$ ) and (B) at pH 2 ( $R^2X = 0.582$ ). In both PCA it is possible to observe the discriminant capability of the volatilome analyzed by HS-SPME/GC-MS, as each cell line forms an independent cluster.

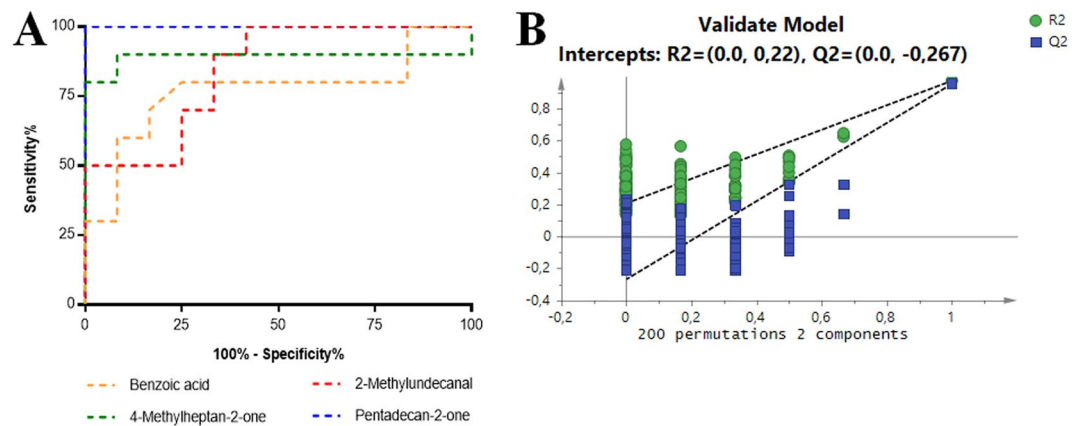
squares discriminant analysis (PLS-DA)) proved that VOCs can discriminate PCa cell lines from normal prostate cell line and between the different PCa cell lines at both pHs as showed in Fig. 2.

To evaluate which VOCs were responsible for this separation, each cancer cell line was compared separately with the normal cell line, using a MVA supervised analysis (PLS-DA), namely 22RV1 vs. PNT2, PC3 vs. PNT2, DU145 vs. PNT2 vs. LNCaP vs. PNT2. An optimal separation between PCa cell lines and normal cell line was observed at both pH values (Supplementary Fig. 2 (pH 7) and Supplementary Fig. 3 (pH 2)).

To prove the robustness of the discrimination, all PLS-DA models were validated through permutation test (200 random permutations of  $Y$ -observations, 2 components). The results of this validation showed that all created models were robust for the discrimination between PCa cell lines and the normal prostate cell line (Fig. 3B and Supplementary Table 1).

All VOCs with Variable Importance to the Projection (VIP) values higher than 1 were considered potentially relevant for the separation among cell lines. Hence, at pH 7 a total of 23 VOCs were considered relevant to differentiate 22RV1 from PNT2 and 32 at pH 2; at pH 7, 16 VOCs were considered relevant to differentiate PC3 from PNT2 and 25 at pH 2; at pH 7, 27 VOCs were considered relevant to differentiate DU145 from PNT2 and 32 at pH 2; and, finally, at pH 7, 21 were considered relevant to differentiate LNCaP from PNT2 and 24 at pH 2 (Tables 1 and 2).

To confirm the importance of these metabolites, univariate analysis was performed as explained below (see statistical analysis section). A total of 8 VOCs proved to be relevant to differentiate 22RV1 from PNT2 in the results obtained at pH 7 and 11 at pH 2; 8 VOCs proved to be relevant to differentiate PC3 from PNT2 at pH 7 and 19 at pH 2; 7 VOCs proved to be relevant to differentiate DU145 from PNT2 at pH 7 and 13 at pH 2; and 7 proved to be relevant to differentiate LNCaP from PNT2 at pH 7 and 13 at pH 2 (Tables 1 and 2). At pH 7, one VOC stood out, namely, pentadecan-2-one since this VOC revealed to be important for the separation among all PCa cell lines and normal cell line (Fig. 3A, Table 1 and Fig. 4). However, several other VOCs were tentatively identified (Supplementary Table 2) and were also able to discriminate between cancer and normal cell lines, namely, 1,3-benzothiazole, 3,7-dimethyloct-7-en-1-ol (22RV1 vs PNT2, PC3 vs PNT2, DU145 vs PNT2), 2-methylundecanal, (PC3 vs PNT2, DU145 vs PNT2 and LNCaP vs PNT2), 2,7-dimethyloctan-1-ol (PC3 vs PNT2), 2-(4-methylcyclohex-3-en-1-yl)propan-2-ol (PC3 vs PNT2), decan-1-ol (DU145 vs PNT2), 1-methoxypropan-2-yl acetate (LNCaP vs PNT2), and 1,4-xylene (LNCaP vs PNT2) (Table 1).



**Figure 3.** (A) Example of assessment of the diagnostic performance through ROC analysis obtained for PC3 vs PNT2 at pH7 (2-pentadecanone (AUC = 1) and 2-methylundecanal (AUC = 0.84)) and at pH2 (4-methylheptan-2-one (AUC = 0.89) and benzoic acid (AUC = 0.77)). (B) Statistical validation of the PLS-DA model for PC3 vs PNT2 at pH 2 by permutation testing (200 permutations; 2 components).

At pH 2, 8 VOCs stood out, namely, cyclohexanone (Fig. 5), 4-methylheptan-2-one, 2-methylpentane-1,3-diol, 4-methylbenzaldehyde (Fig. 5), 1-(3,5-dimethylfuran-2-yl) ethanone, methyl benzoate, nonanoic acid and decanoic acid (Fig. 4) as they revealed to be important for the separation between all PCa cell lines and normal cell line (Table 2). Other VOCs unveiled specificity for each cell line namely, 4-methylpent-3-en-2-one (22RV1 vs PNT2, PC3 vs. PNT2, DU145 vs PNT2), 5-methylheptan-2-one (PC3 vs PNT2, DU145 vs PNT2 and LNCaP vs PNT2), phenylethanol (22RV1 vs PNT2, PC3 vs PNT2, DU145 vs PNT2), 4-methylnonanoic acid (22RV1 vs PNT2 and LNCaP vs PNT2), benzyl acetate (22RV1 vs PNT2 and LNCaP vs PNT2), 5-methyl-2-propan-2-ylcyclohexan-1-ol (PC3 vs PNT2 and LNCaP vs PNT2), 1-ethoxypentane (DU145 vs PNT2 and LNCaP vs PNT2), methyl nonanoate (22RV1 vs PNT2), 2-ethoxy-2-methylbutane (PC3 vs PNT2), hexanoic acid (PC3 vs PNT2), phenylmethanol (PC3 vs PNT2), 2,4-dimethylheptan-1-ol (PC3 vs. PNT2), benzoic acid (PC3 vs PNT2), 6-pentyloxan-2-one (PC3 vs PNT2) (Table 2). Some unidentified VOCs (codified as unknown 1, 2, 3...) were also found as important for the discrimination between cancer and normal cell lines.

ROC curves were performed for all significantly altered VOCs (Fig. 3A), and the results revealed that all significantly altered metabolites have a AUC higher than 0.5 and several metabolites showed a AUC equal to 1 (Tables 1 and 2).

The sensitivity and specificity of the discriminative sets of metabolites obtained for each pair were also calculated and the results reveal that the discriminant sets have a sensitivity and specificity of 100% or very close to this value for discriminating the PCa cell lines from the normal prostate cell line (Supplementary Table 3).

## Discussion

**Comparative analysis of VOCs in PCa and prostate normal cell lines obtained at pH 7.** In this work, we showed the potential of VOCs at physiological pH to discriminate several PCa cell lines from a normal prostate cell line. The results disclosed significantly altered VOCs in all PCa cell lines when compared with the normal cell line (Table 1), with special emphasis for pentadecan-2-one that was found increased in all PCa cell lines. Importantly, this VOC seems to be cell-specific as it is not present in the culture medium, and, therefore, it should be investigated in future as a promising biomarker for PCa.

In addition, our findings further demonstrate that it is possible to discriminate PCa cells with different degrees of aggressiveness based on VOCs profiling. For instance, 2,7-dimethyloctan-1-ol and 2-(4-methylcyclohex-3-en-1-yl)propan-2-ol were found significantly decreased in the culture medium of PC3 cells, which have high metastatic potential (Table 1). The significant increase of decan-1-ol levels was a characteristic alteration of the cell line with moderate metastatic potential (DU145), whereas the significant increase of 3-methylbut-3-en-2-ol, 1-methoxypropan-2-yl acetate and 1,4-xylene levels were characteristic of the cell line with the lower metastatic potential (LNCaP) (Table 1). It was also possible to discriminate androgen responsive from androgen non-responsive cell lines using 3,7-dimethyloct-7-en-1-ol, since this metabolite was only significantly decreased in androgen non-responsive cell lines (PC3, DU145 and 22RV1) (Table 1). Interestingly, 22RV1 that is an androgen receptor positive cell line was able to survive and grow in a culture medium not supplemented with androgen, resulting in the development of an androgen-independent phenotype cell line, possibly by a mechanism similar to the one occurring in clinical practice after hormone deprivation therapy. This explanation was already addressed by other researchers<sup>41</sup>.

The integration of VOCs in metabolic pathways is, at present, hard to perform, because studies with volatiles are scarce and very recent. However, using the human metabolome database (HMDB)<sup>42–44</sup>, it was possible to see if some of those metabolites found in our samples were already found by other researchers in biological matrices. The Supplementary Table 2 enables to do the interconnection between the significantly altered identified metabolites in our *in vitro* disease model and this database<sup>42–44</sup>.

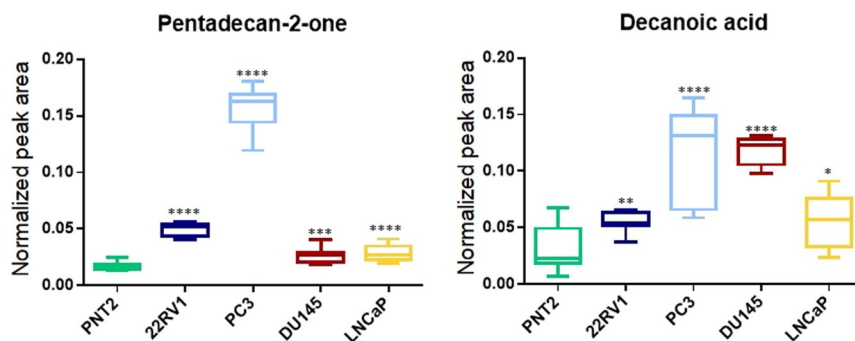
Chemical name (IUPAC) or common name	22RV1 vs PNT2				PC3 vs PNT2				DU145 vs PNT2				LNCaP vs PNT2			
	p value	Variation ± uncertainty	ES ± ES <sub>SE</sub>	AUC	p value	Variation ± uncertainty	ES ± ES <sub>SE</sub>	AUC	p value	Variation ± uncertainty	ES ± ES <sub>SE</sub>	AUC	p value	Variation ± uncertainty	ES ± ES <sub>SE</sub>	AUC
<b>Alcohols</b>																
3-Methylbut-3-en-2-ol		↑	↑										0.0018 <sup>P</sup>	↑34.78 ± 7.88	↑1.48 ± 0.88	0.86
2,7-Dimethyloctan-1-ol (dihydro-citronellol)	>0.05	↓	↓		<0.0001 <sup>P</sup>	↓74.48 ± 7.90	↓5.91 ± 1.84	1	>0.05	↑	↑		>0.05	↓32.61 ± 28.71	↓	
2-(4-Methylcyclohex-3-en-1-yl)propan-2-ol (α-Terpineol)					<0.0001 <sup>P</sup>	↓58.47 ± 4.51	↓7.22 ± 2.18	1								
3,7-Dimethyloct-7-en-1-ol (α-citronellol)	0.0096	↓48.80 ± 20.30	↓1.26 ± 0.85	0.81	<0.0001 <sup>P</sup>	↓58.00 ± 21.18	↓1.52 ± 0.88	0.98	<0.001 <sup>P</sup>	↓73.13 ± 23.15	↓1.86 ± 0.95	1	>0.05	↑137.92 ± 28.06	↑1.14 ± 0.84	
Decan-1-ol									0.0251	↑18.24 ± 6.65	↑1.01 ± 0.84	0.77				
<b>Esters</b>																
1-Methoxypropan-2-yl acetate	>0.05	↓30.40 ± 16.70	↓0.85 ± 0.81						>0.05	↑72.03 ± 8.56	↑1.11 ± 0.85		0.0447	↑70.57 ± 18.49	↑1.11 ± 0.83	0.74
<b>Ketones</b>																
Pentadecan-2-one	<0.0001 <sup>P</sup>	↑203.50 ± 5.40	↑7.34 ± 2.22	1	<0.0001 <sup>P</sup>	↑848.29 ± 6.06	↑10.53 ± 3.07	1	0.0007 <sup>P</sup>	↑58.66 ± 10.04	↑1.89 ± 0.96	0.94	<0.0001 <sup>P</sup>	↑73.12 ± 9.26	↑2.27 ± 1.01	0.98
<b>Aromatic hydrocarbon like</b>																
1,4-Xylene					>0.05	↓10.74 ± 6.87	↓		>0.05	↑89.15 ± 32.29	↑		0.0011 <sup>P</sup>	↑182.61 ± 24.11	↑1.56 ± 0.89	0.88
Naphthalene	0.0011 <sup>P</sup>	↑49.20 ± 8.80	↑1.76 ± 0.92	0.88												
<b>Aldehydes</b>																
2-Methylundecanal					0.0029 <sup>P</sup>	↑58.67 ± 11.00	↑1.54 ± 0.89	0.84	0.0037	↓31.88 ± 10.31	↓1.41 ± 0.88	0.84		↑71.77 ± 15.15	↑1.37 ± 0.86	0.82
<b>Others</b>																
1,3-Benzothiazole	<0.0001 <sup>P</sup>	↓75.00 ± 6.80	↓6.91 ± 2.10	1	<0.0001 <sup>P</sup>	↓70.22 ± 7.35	↓7.03 ± 2.13	1	0.0348	↓23.35 ± 8.56	↓1.02 ± 0.84	0.70	<0.0001 <sup>P</sup>	↓69.91 ± 7.11	↓5.96 ± 1.86	1
<b>Unknowns</b>																
Unknown 2									0.0414	↓32.26 ± 15.95	↓0.92 ± 0.83	0.71				
Unknown 3									>0.05	↓20.36 ± 12.46	↓		0.0204	↑30.56 ± 10.07	↑1.04 ± 0.83	0.78
Unknown 7	0.0011 <sup>P</sup>	↓58.30 ± 22.20	↓1.46 ± 0.88	0.88	>0.05	↓	↓		0.0010 <sup>P</sup>	↓23.26 ± 6.27	↓1.63 ± 0.91	0.85	0.0023 <sup>P</sup>	↓34.10 ± 15.37	↓1.05 ± 0.83	0.85
Unknown 8	0.0024	↑14.60 ± 3.80	↑1.43 ± 0.87	0.84					>0.05	↓9.18 ± 5.37	↓		>0.05	↑	↑	
Unknown 12									0.0338	↑20.39 ± 7.80	↑0.96 ± 0.84	0.76				
Unknown 13	<0.0001 <sup>P</sup>	↑63.50 ± 4.70	↑4.03 ± 1.38	1	<0.0001 <sup>P</sup>	↑177.78 ± 6.14	↑6.03 ± 1.87	1	>0.05	↑	↑		0.0016 <sup>P</sup>	↑36.55 ± 7.49	↑1.63 ± 0.90	0.85
Unknown 14	0.0003 <sup>P</sup>	↑58.60 ± 10.10	↑1.76 ± 0.92	0.91	<0.0001 <sup>P</sup>	↑181.87 ± 4.63	↑8.10 ± 2.42	1	>0.05	↓	↓					

**Table 1.** List of VOCs, from the analysis of samples prepared at pH 7, selected as important in discriminating PCa cell lines from normal cell lines (22RV1 vs PNT2, PC3 vs PNT2, DU145 vs PNT2 and LNCaP vs PNT2). VOCs are characterized by their IUPAC name. Values for *p*-values percentage of variation, effect size (ES), standard error (ES<sub>SE</sub>) and area under the curve (AUC) are represented for each VOC. <sup>P</sup>Alterations remaining significant after Bonferroni correction, with cutoff *p* value of:  $2.17 \times 10^{-3}$  (0.05 divided by 23 analyzed VOCs) for 22RV1 vs PNT2;  $3.12 \times 10^{-3}$  (0.05 divided by 16 analyzed VOCs) for PC3 vs PNT2;  $1.85 \times 10^{-3}$  (0.05 divided by 27 analyzed VOCs) for DU145 vs PNT2; and  $2.38 \times 10^{-3}$  (0.05 divided by 21 analyzed VOCs) for LNCaP vs PNT2.

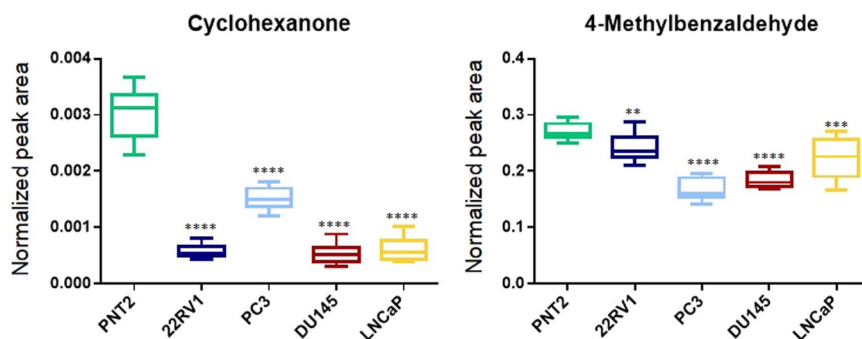
The increase in pentadecan-2-one observed in all PCa cell lines is in accordance with previously reported increase in ketone levels in urine from PCa patients<sup>38</sup>, as well as in other cancer cell lines<sup>37,39,45,46</sup>. This finding may be explained by the increased fatty acid  $\beta$ -oxidation and protein metabolism, used to produce energy, which lead to ketones production<sup>47</sup>. In particular, the metabolism of C15:0 long-chain fatty acids may originate pentadecan-2-one in PCa cell lines<sup>48</sup>. Furthermore, ketones can also be produced from the respective secondary alcohols<sup>37</sup>. Alteration in the activity of alcohol dehydrogenase<sup>49</sup> and cytochrome P450 (CYP 450) activities<sup>30</sup> can also be related to alteration in ketone and aldehydes levels, since hydrocarbons can be metabolized to aldehydes or ketones in the human body via these enzymes<sup>50</sup>. In our study, alterations in the aldehydes and hydrocarbons levels were also observed, which reinforce the connection between the observed changes and alterations in the activity of these two enzymes.

Chemical name (IUPAC) or common name	22RV1 vs PNT2				PC3 vs PNT2				DU145 vs PNT2				LNCaP vs PNT2			
	p value	Variation ± uncertainty	ES ± ES <sub>SE</sub>	AUC	p value	Variation ± uncertainty	ES ± ES <sub>SE</sub>	AUC	p value	Variation ± uncertainty	ES ± ES <sub>SE</sub>	AUC	p value	Variation ± uncertainty	ES ± ES <sub>SE</sub>	AUC
<b>Alcohols</b>																
2-Methylpentane-1,3-diol	0.0105	↓10.70 ± 3.80	↓1.17 ± 0.84	0.81	<0.0001 <sup>P</sup>	↓45.10 ± 5.81	↓3.95 ± 1.36	1	<0.0001 <sup>P</sup>	↓25.79 ± 3.95	↓2.95 ± 1.14	0.99	<0.0001 <sup>P</sup>	↓24.26 ± 5.53	↓1.97 ± 0.95	0.92
Phenyl-methanol	>0.05	↑35.50 ± 19.60	↑		0.0189	↓33.69 ± 15.30	↓1.04 ± 0.83	0.82	>0.05	↑	↑		>0.05	↑	↑	
2,4-Dimethylheptan-1-ol	>0.05	↑32.70 ± 20.10	↑		0.0210	↓35.32 ± 16.50	↓1.02 ± 0.83	0.79	>0.05	↓	↓82.41 ± 7.28		>0.05	↑	↑	
Phenylethanol	<0.0001 <sup>P</sup>	↓38.10 ± 7.90	↓2.36 ± 1.02	0.97	<0.0001 <sup>P</sup>	↓69.81 ± 8.12	↓5.20 ± 1.66	1	<0.0001 <sup>P</sup>	↓38.79 ± 8.76	↓2.16 ± 0.99	0.92				
5-Methyl-2-propan-2-ylcyclohexan-1-ol (DL-menthol)	>0.05	↓	↓		0.0068	↓45.82 ± 18.48	↓1.27 ± 0.85	0.83	0.0447	↓27.4 ± 14.89	↓0.84 ± 0.81	0.74	0.0019 <sup>P</sup>	↓42.10 ± 13.39	↓1.57 ± 0.89	0.91
<b>Esters</b>																
Methyl benzoate	<0.0001 <sup>P</sup>	↓42.20 ± 3.90	↓5.51 ± 1.74	1	<0.0001 <sup>P</sup>	↓53.12 ± 4.10	↓6.95 ± 2.11	1	<0.0001 <sup>P</sup>	↓51.06 ± 5.58	↓4.84 ± 1.57	1	<0.0001 <sup>P</sup>	↓22.45 ± 4.27	↓2.33 ± 1.01	0.94
Benzyl acetate	<0.0001 <sup>P</sup>	↓35.17 ± 6.10	↓2.75 ± 1.10	1									0.0090	↑23.73 ± 7.0	↑1.19 ± 0.84	0.79
Methyl nonanoate	<0.0001 <sup>P</sup>	↑366.00 ± 7.90	↑6.46 ± 1.98	1												
<b>Ketones</b>																
4-Methylpent-3-en-2-one	<0.0001 <sup>P</sup>	↓47.20 ± 7.10	↓3.44 ± 1.24	0.99	<0.0001 <sup>P</sup>	↓62.38 ± 7.88	↓4.54 ± 1.50	1	<0.0001 <sup>P</sup>	↓48.54 ± 9.03	↓2.80 ± 1.11	0.95				
Cyclohexanone	<0.0001 <sup>P</sup>	↓81.10 ± 7.00	↓7.72 ± 2.32	1	<0.0001 <sup>P</sup>	↓50.48 ± 5.88	↓4.53 ± 1.50	1	<0.0001 <sup>P</sup>	↓82.41 ± 7.28	↓7.59 ± 2.28	1	<0.0001 <sup>P</sup>	↓80.15 ± 7.33	↓7.19 ± 2.18	1
4-Methylheptan-2-one	<0.0001 <sup>P</sup>	↓28.60 ± 6.30	↓2.10 ± 0.97	0.92	0.0003 <sup>P</sup>	↓38.37 ± 8.84	↓2.11 ± 0.98	0.89	<0.0001 <sup>P</sup>	↓36.06 ± 6.93	↓2.51 ± 1.05	0.95	0.0014 <sup>P</sup>	↓26.30 ± 7.86	↓1.52 ± 0.88	0.84
5-Methylheptan-2-one					<0.0001 <sup>P</sup>	↓51.58 ± 6.11	↓4.48 ± 1.48	1	<0.0001 <sup>P</sup>	↓27.47 ± 5.97	↓2.10 ± 0.97	0.94	0.0014 <sup>P</sup>	↓20.36 ± 5.95	↓1.50 ± 0.88	0.86
1-(3,5-Dimethylfuran-2-yl) ethanone	<0.0001 <sup>P</sup>	↓47.80 ± 5.30	↓4.79 ± 1.56	1	<0.0001 <sup>P</sup>	↓41.51 ± 5.09	↓4.06 ± 1.38	1	<0.0001 <sup>P</sup>	↓46.45 ± 6.01	↓3.96 ± 1.36	1	0.0003 <sup>P</sup>	21.78 ± 5.50	↓1.75 ± 0.91	0.88
6-Pentylloxan-2-one (δ-decalactone)	>0.05	↓	↓		0.0006 <sup>P</sup>	↓39.93 ± 11.98	↓1.64 ± 0.90	0.88	>0.05	↓	↓		>0.05	↑	↑	
<b>Alkane hydrocarbon</b>																
2-Ethoxy-2-methylbutane	>0.05	↓26.60 ± 19.90	↓		0.0342	↓38.26 ± 19.77	↓0.94 ± 0.81	0.74					>0.05	↓	↓	
1-Ethoxy-pentane					0.0056	↑	↑		0.0010 <sup>P</sup>	↓31.16 ± 9.21	↓1.58 ± 0.89	0.88	0.0001 <sup>P</sup>	↓35.34 ± 8.69	↓1.95 ± 0.95	0.92
<b>Aldehydes</b>																
4-Methyl-benzaldehyde	0.0022	↓10.60 ± 3.00	↓1.47 ± 0.88	0.81	<0.0001 <sup>P</sup>	↓38.00 ± 2.98	↓6.20 ± 1.92	1	<0.0001 <sup>P</sup>	↓32.04 ± 2.39	↓6.29 ± 1.94	1	0.0005 <sup>P</sup>	↓18.39 ± 4.31	↓1.85 ± 0.93	0.87
<b>Organic acids</b>																
Hexanoic acid	>0.05	↓	↓		0.0092	↓40.00 ± 16.66	↓1.18 ± 0.84	0.84	>0.05	↓	↓		>0.05	↑	↑	
Benzoic acid	>0.05	↓	↓		0.0331	↓37.17 ± 17.66	↓1.01 ± 0.82	0.77	>0.05	↓	↓		>0.05	↑	↑	
Nonanoic acid	<0.0001 <sup>P</sup>	↓55.60 ± 8.80	3.47 ± 1.25	1	<0.0001 <sup>P</sup>	59.85 ± 7.41	4.54 ± 1.50	1	<0.0001 <sup>P</sup>	↑31.32 ± 6.29	↑2.33 ± 1.02	0.97	<0.0001 <sup>P</sup>	↑66.28 ± 7.67	↑5.09 ± 1.64	1
4-Methyl-nonanoic acid	<0.0001 <sup>P</sup>	↑1217.10 ± 11.70	↑5.79 ± 1.81	1									<0.0001 <sup>P</sup>	↑3242.3 ± 10.90	↑65.79 ± 2.07	1
Decanoic acid	0.0027	↑71.10 ± 14.00	↑1.47 ± 0.88	0.83	<0.0001 <sup>P</sup>	↑250.44 ± 18.10	↑2.42 ± 1.03	0.98	<0.0001 <sup>P</sup>	↑268.37 ± 8.66	↑5.21 ± 1.67	1	0.0122	74.50 ± 19.00	↑1.13 ± 0.83	0.80
<b>Unknowns</b>																
Unknown 16	>0.05	↓	↓		0.0321	↓30.20 ± 14.74	↓0.95 ± 0.82	0.79								
Unknown 18	<0.0001 <sup>P</sup>	↓94.90 ± 11.80	↓6.05 ± 1.88	1	<0.0001 <sup>P</sup>	↓59.85 ± 7.41	↓4.54 ± 1.50	1								
Unknown 19	>0.05	↓20.90 ± 17.10	↓		0.0004 <sup>P</sup>	↓38.38 ± 10.98	↓1.71 ± 0.91	0.92	>0.05	↓	↓		>0.05	↑	↑	

**Table 2.** List of VOCs, from the analysis of samples prepared at pH 2, selected as important in discriminating PCA cell lines from normal cell lines (22RV1 vs PNT2, PC3 vs PNT2, DU145 vs PNT2 and LNCaP vs PNT2). VOCs are characterized by their IUPAC name. Values for *p*-values, percentage of variation, effect size (ES), standard error (ES<sub>SE</sub>), and area under the curve (AUC) are represented for each VOC. ↑, VOCs increased; ↓, VOCs decreased in the extracellular medium of PCA compared with normal cell line <sup>P</sup>Alterations remaining significant after Bonferroni correction, with cutoff *p* value of  $1.56 \times 10^{-3}$  (0.05 divided by 32 analyzed VOCs) for 22RV1 vs PNT2;  $2.00 \times 10^{-3}$  (0.05 divided by 25 analyzed VOCs) for PC3 vs PNT2;  $2.38 \times 10^{-3}$  (0.05 divided by 21 analyzed VOCs) for DU145 vs PNT2; and  $2.08 \times 10^{-3}$  (0.05 divided by 24 analyzed VOCs) for LNCaP vs PNT2.



**Figure 4.** Boxplots from the metabolite pentadecan-2-one, increased in all PCa cells when compared with PNT2 (normal cells), after univariate analysis, obtained at pH 7 and boxplots from the metabolite decanoic acid, increased in all PCa cells when compared with PNT2 (normal cells), after univariate analysis, obtained at pH 2.



**Figure 5.** Boxplots from cyclohexanone and 4-methylbenzaldehyde, metabolites decreased in all PCa cells when compared with PNT2 (normal cells), after univariate analysis, obtained at pH 2.

The alteration in aldehydes (e.g. 2-methylundecanal) profiling is common in urine of PCa patients<sup>38</sup> and in cancer cell lines from colon and lung<sup>37,39,51</sup>. This altered profile may be associated with an alteration in lipid peroxidation, induced by the increased ROS levels, characteristic of cancer cell metabolism and inflammation<sup>38,47,51,52</sup>. Beyond lipid peroxidation, aldehydes can also result from amino acid and carbohydrate catabolism<sup>53,54</sup>. An alternative explanation for the aldehyde levels variation can be related to alterations in aldehyde dehydrogenases (ALDHs) activity, responsible for the aldehydes oxidation to carboxylic acids<sup>53,54</sup>. Furthermore, ALDHs are involved in cellular proliferation, differentiation and survival, and in the cellular response to oxidative stress. The alteration in ALDHs was associated with PCa and other cancers progression<sup>53,54</sup>.

A significant alteration in aromatic hydrocarbons, such as naphthalene and 1,4-xylene, in the culture media of cancer cell lines was also observed. Importantly, the alteration in 1,4-xylene levels observed in our study is in concordance with previously reported alteration in 1,4-xylene in exhaled breath from PCa patients<sup>55</sup> and in urines from head and neck cancer patients<sup>56</sup>. The presence of these molecules was previously related with the presence of ROS and oxidative stress<sup>47,57,58</sup>. So, the increase in aromatic hydrocarbons in cancer cell lines media can be an indicator of an increase in ROS activity, which could lead to lipid peroxidation and consequently to cell membrane damage<sup>59</sup>.

In our study, several alcohols, such as 3-methylbut-3-en-2-ol, 2,7-dimethyloctan-1-ol, 2-(4-methylcyclohex-3-en-1-yl)propan-2-ol, 3,7-dimethyloct-7-en-1-ol and decan-1-ol were altered in PCa cell lines; this alteration in alcohols levels may be due to the alteration in hydrocarbon metabolism, being alcohols end-products of this metabolism<sup>50,58</sup>. The  $\beta$ -oxidation used by cancer cells to produce acetyl-CoA, that can be incorporated in Krebs cycle, may be other important source of alcohols. Thus, one may hypothesize that decan-1-ol can be associated with the synthesis of a medium-chain fatty acid, namely capric acid (C10:0), for energy production<sup>42–44</sup>. Other well described characteristic of cancer cells is their rapid growth, which implies an increase of cellular membrane synthesis. The alcohols may be metabolized to carboxylic acids and these acids further used to the synthesis of cellular membrane precursors<sup>60</sup>. Another possible explanation for the increased levels of some alcohols is the induction of the cytochrome P450 enzymes that occurs during carcinogenesis<sup>56</sup>. CYP 450 can hydroxylate several VOCs, including the alkanes produced during lipid peroxidation of polyunsaturated fatty acids, which leads to the production of corresponding alcohols<sup>56</sup>. Alcohols can also be produced by enzymatic reduction of aldehydes by alcohol dehydrogenases<sup>37</sup>.

**Comparative analysis of VOCs in PCa and normal cell lines obtained at pH 2.** It is well established that tumor microenvironment is acidic<sup>40</sup>, to promote tumor progression and metastasis. The biochemical mechanisms related with acidic pH include hypoxia, excessive glycolysis, hyperexpression of carbonic anhydrase and

poor perfusion<sup>40</sup>. The important role of acidic pH in tumor microenvironment highlights the interest of studying the volatilome at pH 2.

As for physiological pH, the volatilome obtained at pH 2 was able to differentiate PCa cell lines from normal prostate cell line, taking into account the quality of the two PLS-DA models ( $R^2X = 0.526$ ;  $R^2Y = 0.457$ ;  $Q^2 = 0.437$  for pH 2 vs  $R^2X = 0.445$ ;  $R^2Y = 0.463$ ;  $Q^2 = 0.446$  for pH 7). Although the discriminant capability of these two models is similar, the discriminant metabolites were not the same at the two medium pH, showing that acidification of the samples markedly influences the prostate cell volatilome (e.g. detection of volatile organic acids). As described by other authors<sup>30</sup>, the detected volatilome was markedly different in the two pHs in study, which highlights the importance of the study of volatilome at different pH, to obtain a comprehensive picture of cells volatilome.

Seven VOCs, namely 2-methylpentane-1,3-diol, 1-(3,5-dimethylfuran-2-yl)ethanone, methyl benzoate, nonanoic acid, cyclohexanone, 4-methylbenzaldehyde (Fig. 5), and 4-methylheptan-2-one were significantly decreased in the extracellular medium of all PCa cell lines when compared with normal cell line (Table 2). On the other hand, decanoic acid (Fig. 4) was increased in the extracellular medium of all PCa cell lines. Based on the altered volatile profile among cell lines in study, it was also possible to discriminate PCa cells with different aggressiveness. For example, 2-ethoxy-2-methylbutane acid, hexanoic acid, phenylmethanol, 2,4-dimethylheptan-1-ol, benzoic acid, and 6-pentyloxan-2-one were significantly decreased in the extracellular medium of high metastatic potential cell line (PC3), whereas 1-ethoxypentane was significantly decreased specifically in moderate and low metastatic potential cell line (DU145 and LNCaP, respectively). Moreover, some VOCs can discriminate androgen responsive cell lines from androgen non-responsive cell lines. For instance, 4-methylpent-3-en-2-one and phenylethanol (Table 2) were only significantly decreased in the androgen non-responsive cell lines (PC3, DU145 and 22RV1). Importantly, 4-methylnonanoic acid, hexanoic acid and 6-pentyloxan-2-one were absent in the cellular medium, suggesting its origin from cellular metabolism, thus unveiling great potential as candidate biomarkers to be further validated in biofluids from PCa patients.

The majority of the VOCs significantly altered were decreased in cancer cell lines in comparison with normal cell line, being this observation previously made in other metabolomic studies with urine from PCa patients<sup>38,61</sup>. The cancer cells may use these metabolites for their metabolic processes more extensively than normal prostate cells<sup>38,61</sup>, converting these VOCs in other metabolites that cannot be detected by our methodology.

Comparison of the metabolites found in our samples at pH 2 was made with those already described by other researchers in biological matrices<sup>42–44</sup> (Supplementary Table 2). The significant alterations in the levels of decanoic acid in the PCa cells study is in concordance with previously reported in serum of PCa patients<sup>62</sup>. Also, alterations in decanoic acid, phenylmethanol and nonanoic acid were associated with melanoma<sup>63</sup>. The significant alteration in the levels of 4-methylheptan-2-one study is in concordance with previously reported in urines from head and neck cancer patients<sup>56</sup>.

Similarly to pH 7, in the volatile profile obtained at pH 2, it was observed a significant alteration in the levels of ketones, aldehydes and alcohols associated with PCa. Significant alterations in levels of some esters and organic acids were also observed.

Esters can be found naturally in fats and may be originated from acids and alcohols during lipid hydrolysis, which may explain the reduction in the levels of these VOCs, observed in PCa cell lines<sup>50</sup>. So, the alteration in esters levels and organic acids levels can be related as ester metabolism can lead to organic acid production and acid metabolism can lead esters formation<sup>37</sup>.

Alteration in organic acids was previously described in urine and serum of PCa patients<sup>61,62</sup>. Organic acids can be involved in several biological processes, including cell signaling, energy storage, energy source, and cellular membrane integrity<sup>61</sup>. As aforementioned, alteration in ALDHs function can also be associated with the alteration in the organic acid levels, as this enzyme can catalyze the conversion of aldehydes into organic acids<sup>37</sup>.

## Conclusion

Our results highlight the capability of volatilome analysis to identify potential biomarkers to be used in PCa diagnosis, since at both analyzed pH the VOCs profile was able of differentiate all cell lines in study. One of the most problematic drawbacks of PSA is its inability to differentiate aggressive PCa from indolent PCa. Our results reveal that through the evaluation of the alterations in volatile profile of PCa cell lines it is possible to differentiate PCa cells with different aggressiveness. The use in clinical practice of a biomarker or a panel of biomarkers able to make this differentiation will bring major benefits for PCa patients, improving PCa management and circumventing the problem of overtreatment. Furthermore, volatilome also prove be able to differentiate PCa cell lines with different hormonal dependency.

The influence of pH in the volatilome was also evaluated and is a parameter to be considered. Acidification of the samples markedly influenced the volatile composition (e.g. detection of volatile organic acids), hence, to have a more comprehensive analysis of cell volatilome it is advantageous to use both pH.

The integration of VOCs in specific metabolic pathways is still very difficult, needing more studies to completely accomplish this goal. Nevertheless, our results indicate that PCa cell lines have lipid metabolism altered, mainly in lipid peroxidation, lipid hydrolysis, synthesis of fatty acid and fatty acid  $\beta$ -oxidation. The obtained results also suggest the modification in the activity of some important enzymes, namely, alcohol dehydrogenase, cytochrome P450 and aldehyde dehydrogenase.

It is important to emphasize that for a possible use in the clinical practice it is mandatory to prove the translatability of these results to *in vivo* samples, preferentially urine. Further studies are still needed to confirm our results, and find out a non-invasive, sensible and specific biomarker (or set of biomarkers) for prostate cancer.



## Materials and Methods

**Chemicals.** All chemicals and reagents were of analytical grade. RPMI-1640 medium, 4-fluorobenzaldehyde (internal standard) and all GC-MS standard compounds were purchased from Sigma-Aldrich Co. (St. Louis, MO, USA). The antibiotic mixture penicillin/streptomycin (10000 U/mL/10000 µg/mL), heat inactivated fetal bovine serum (FBS) and trypsin 0.25%-EDTA were purchased from GIBCO Invitrogen (Barcelona, Spain). Hydrochloric acid (HCl) and sodium hydrogen carbonate were obtained from Merck (Darmstadt, Germany). Sodium chloride (NaCl) was from VWR (Leuven, Belgium).

**Cell Culture.** PCa human immortalized cell lines (PC3, 22RV1, DU145 and LNCaP) and human normal prostate epithelium immortalized cell line (PNT2) were provided by Portuguese Oncology Institute-Porto (IPO-Porto) and their respective characteristics are shown in Supplementary Table 4. In brief, 22RV1 and LNCaP are cell lines from prostate carcinoma with androgen receptor (AR) expression, whereas PC3 and DU145 cells lines are androgen unresponsive. PC3 is a cell line from grade IV adenocarcinoma with high metastatic potential; DU145 is a cell line from prostate carcinoma, with moderated metastatic potential; and LNCaP is a cell line from prostate carcinoma with low metastatic potential. The selection of the cells lines was performed to allow to cover an extended range of PCa with different features, namely different aggressiveness and hormonal dependency. Furthermore, it was possible to use the same culture media to grow all cell lines. The use of different culture media, in this kind of studies, may be a confounding factor, once different culture media might have different effects on cellular metabolic profiles, thereby compromising the comparison of the VOCs profiles<sup>64</sup>.

All cell lines were grown in RPMI-1640 supplemented with 10% of FBS and 1% of penicillin/streptomycin, maintained at 37 °C and 5% CO<sub>2</sub> in 75 cm<sup>2</sup> culture flasks and grown to 80% confluency, before each passage. For all cell lines the same passages were used. All cell lines were routinely tested for Mycoplasma spp. contamination (PCR Mycoplasma Detection Set, Clontech Laboratories).

**VOCs collection and extraction.** The experiments were carried out during 4 consecutive passages (passages 3 to 6) with triplicates for each passage, after an adaptation stage of at least 3 passages for all cell cultures. After the cells had reached near confluence, the medium was discarded and replaced with 15 mL of fresh RPMI-1640 medium, and incubated at 37 °C and 5% CO<sub>2</sub> for 48 h, together with three blanks (cellular medium without cells treated exactly as the study samples). After 48 h, the extracellular medium from flasks with cells and blanks were collected, centrifuged (2000 × g for 10 min at 4 °C), the supernatant separated from the pellet, and divided into two equal aliquots (one for analysis at pH 7 and another for analysis at pH 2). To prevent the loss of VOCs, all collection steps were performed on ice and immediately frozen at −80 °C until analysis.

Stored samples were thawed slowly at low temperatures, in order to minimize the loss of VOCs<sup>65</sup>. All samples were analyzed at the pH of the medium culture around pH 7 (pH = 7.675 ± 0.28) and at pH 2 (pH = 2.156 ± 0.31). For acidification, a fixed volume of 5 M HCl was added. For GC-MS analysis, 2 mL of sample were placed into a 10 mL glass vial with the internal standard (4-fluorobenzaldehyde (10 µg/mL)) and NaCl (≈0.59 g). The extraction was performed according to the conditions previously determined<sup>65</sup>. Simultaneously, quality control (QCs) samples were also prepared using the same protocol. QCs samples were a pool of all samples and blanks used in the study. To avoid the frequent freezing and thawing, that could change the VOCs fingerprint, the QCs samples were divided in several aliquots and each aliquot was defrost prior to GC-MS analysis.

**GC-MS analysis.** The data acquisition was performed in a Scion-436 gas chromatograph coupled to a Bruker SQ (single quadrupole) equipped with a SCION SQ ion trap mass detector and a Bruker Daltonics MS workstation software version 6.8, with a Rxi-5Sil MS (30 m × 0.25 mm × 0.25 µm) column from RESTEK. A CombiPAL automatic autosampler (Varian, Palo Alto, CA) was used and experimental conditions were previously described<sup>65</sup>. The carrier gas used was helium C-60 (Gasin, Portugal) (flow of 1 mL/min) and the injector port was heated to 230 °C. The analysis was performed in Full Scan mode. The oven temperature was fixed at 40 °C for 1 min, then increasing to 250 °C (rate 5 °C/min), held for 5 min, then increasing to 300 °C (rate 5 °C/min) and held for 1 min. The transfer line temperature was 280 °C, manifold temperature was 50 °C and the trap temperature was 180 °C. The mass range was 40–350 m/z, with a scan rate of 6 scan/s. All samples were injected randomly<sup>65</sup>.

To ensure reproducibility, QCs samples were injected at the same conditions on every 6 samples (three times per day)<sup>66</sup>.

The identification of VOCs was accomplished by using standards, the National Institute of Standards and Technology (NIST 14) data base spectra library and by comparing experimental Kovats index and Kovats index from literature (Supplementary Table 2).

**Statistical analysis.** Prior to statistical analysis of results, all chromatograms were pre-processed. This pre-processing includes: baseline correction, peak detection, chromatogram deconvolution and alignment, performed using the program MZmine<sup>67</sup>. The parameters used to accomplish these were: RT range 2.8–34.0 min; m/z range 50–250; MS data noise level  $1.0 \times 10^4$ ; m/z tolerance 0.5; chromatogram baseline level  $1.0 \times 10^3$ ; peak duration range 0.02–0.30 min. Also, all ions with a relative standard deviation<sup>68</sup> greater than 30% as well as ions (m/z) coming from the column, the fiber, among others were manually removed from the matrix. The obtained data was normalized for the total area of the chromatograms.

The statistical treatment includes unsupervised (PCA) and supervised analyzes (PLS-DA). All VOCs with VIP values greater than one were considered potential relevant for the separation among cell lines. For these relevant compounds, univariate analysis was performed, using Shapiro-Wilk test, to determine normality distribution of data, and unpaired Student's t-test with Welch correction test, for normal distribution, or unpaired Mann-Whitney test, for non-normal distribution, to calculate the *p* value. In addition, the percentage of variation,

uncertainty of the percentage of variation, and effect size and the standard error were calculated<sup>69</sup>. Bonferroni correction was used to adjust *p*-values. For all significantly (*p* value < 0.05, effect size higher than the standard error and % variation superior to uncertainty) altered metabolites, receiver operating characteristic<sup>10</sup> curves were computed and AUC (area under the curve) was also calculated (using MetaboAnalyst 3.0).

Finally, to confirm the robustness of the PLS-DA models, all PLS-DA models were validated through permutation test (200 random permutations of *Y*-observations, 2 components). The sensitivity and specificity of the discriminant sets of metabolites obtained from all comparison pairs at pH 7 and pH 2 were also calculated (using MetaboAnalyst 3.0).

## References

- Kelly, R. S., Vander Heiden, M. G., Giovannucci, E. & Mucci, L. A. Metabolomic Biomarkers of Prostate Cancer: Prediction, Diagnosis, Progression, Prognosis, and Recurrence. *Cancer Epidemiol Biomarkers Prev* **25**, 887–906, <https://doi.org/10.1158/1055-9965.EPI-15-1223> (2016).
- Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics. *CA Cancer J Clin* **67**, 7–30, <https://doi.org/10.3322/caac.21387> (2017).
- Gaudreau, P. O., Stagg, J., Soulieres, D. & Saad, F. The Present and Future of Biomarkers in Prostate Cancer: Proteomics, Genomics, and Immunology Advancements. *Biomark Cancer* **8**, 15–33, <https://doi.org/10.4137/BIC.S31802> (2016).
- Catalona, W. J. *et al.* Comparison of percent free PSA, PSA density, and age-specific PSA cutoffs for prostate cancer detection and staging. *Urology* **56**, 255–260 (2000).
- Catalona, W. J., Smith, D. S. & Ornstein, D. K. Prostate cancer detection in men with serum PSA concentrations of 2.6 to 4.0 ng/mL and benign prostate examination. Enhancement of specificity with free PSA measurements. *JAMA* **277**, 1452–1455 (1997).
- Sokoloff, M. H., Yang, X. J., Fumo, M., Mhoon, D. & Brendler, C. B. Characterizing prostatic adenocarcinomas in men with a serum prostate specific antigen level of 4.0 ng/mL. *BJU Int* **93**, 499–502 (2004).
- Nadler, R. B. *et al.* Use of 2.6 ng/ml prostate specific antigen prompt for biopsy in men older than 60 years. *The Journal of urology* **174**, 2154–2157, discussion 2157 (2005).
- Lucia, M. S. *et al.* Pathologic characteristics of cancers detected in The Prostate Cancer Prevention Trial: implications for prostate cancer detection and chemoprevention. *Cancer Prev Res (Phila)* **1**, 167–173, <https://doi.org/10.1158/1940-6207.CAPR-08-0078> (2008).
- Thompson, I. M. *et al.* The influence of finasteride on the development of prostate cancer. *N Engl J Med* **349**, 215–224, <https://doi.org/10.1056/NEJMoa030660> (2003).
- Trock, B. J. Application of metabolomics to prostate cancer. *Urologic oncology* **29**, 572–581, <https://doi.org/10.1016/j.urolonc.2011.08.002> (2011).
- Roberts, M. J., Schirra, H. J., Lavin, M. F. & Gardiner, R. A. Metabolomics: a novel approach to early and noninvasive prostate cancer detection. *Korean J Urol* **52**, 79–89, <https://doi.org/10.4111/kju.2011.52.2.79> (2011).
- Andriole, G. L. *et al.* Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* **360**, 1310–1319, <https://doi.org/10.1056/NEJMoa0810696> (2009).
- Schroder, F. H. *et al.* Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* **360**, 1320–1328, <https://doi.org/10.1056/NEJMoa0810084> (2009).
- Heijnsdijk, E. A. *et al.* Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *British journal of cancer* **101**, 1833–1838, <https://doi.org/10.1038/sj.bjc.6605422> (2009).
- Moyer, V. A. & Force, U. S. P. S. T. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* **157**, 120–134, <https://doi.org/10.7326/0003-4819-157-2-201207170-00459> (2012).
- Bostrom, P. J. *et al.* Genomic Predictors of Outcome in Prostate Cancer. *Eur Urol* **68**, 1033–1044, <https://doi.org/10.1016/j.eururo.2015.04.008> (2015).
- Aboud, O. A. & Weiss, R. H. New opportunities from the cancer metabolome. *Clinical chemistry* **59**, 138–146, <https://doi.org/10.1373/clinchem.2012.184598> (2013).
- Monteiro, M. S., Carvalho, M., Bastos, M. d. L. & Pinho, P. G. In *Identification and Data Processing Methods in Metabolomics Future Science Book Series*, 166–184 (FutureScience Ltd, 2015).
- Cuperlovic-Culf, M., Barnett, D. A., Culf, A. S. & Chute, I. Cell culture metabolomics: applications and future directions. *Drug discovery today* **15**, 610–621, <https://doi.org/10.1016/j.drudis.2010.06.012> (2010).
- Halama, A. Metabolomics in cell culture—a strategy to study crucial metabolic pathways in cancer development and the response to treatment. *Archives of biochemistry and biophysics* **564**, 100–109, <https://doi.org/10.1016/j.abb.2014.09.002> (2014).
- Zhang, A., Sun, H., Xu, H., Qiu, S. & Wang, X. Cell metabolomics. *Omic: a journal of integrative biology* **17**, 495–501, <https://doi.org/10.1089/omi.2012.0090> (2013).
- Keshari, K. R. *et al.* Metabolic reprogramming and validation of hyperpolarized <sup>13</sup>C lactate as a prostate cancer biomarker using a human prostate tissue slice culture bioreactor. *Prostate* **73**, 1171–1181, <https://doi.org/10.1002/pros.22665> (2013).
- Monteiro, M. *et al.* GC-MS metabolomics-based approach for the identification of a potential VOC-biomarker panel in the urine of renal cell carcinoma patients. *J Cell Mol Med*, <https://doi.org/10.1111/jcmm.13132> (2017).
- Schmidt, K. & Podmore, I. Current Challenges in Volatile Organic Compounds Analysis as Potential Biomarkers of Cancer. *Journal of biomarkers* **2015**, 981458, <https://doi.org/10.1155/2015/981458> (2015).
- Shirasu, M. & Touhara, K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. *Journal of biochemistry* **150**, 257–266, <https://doi.org/10.1093/jb/mvr090> (2011).
- Di Lena, M., Porcelli, F. & Altomare, D. F. Volatile organic compounds as new biomarkers for colorectal cancer: a review. *Colorectal Dis* **18**, 654–663, <https://doi.org/10.1111/codi.13271> (2016).
- Kell, D. B. *et al.* Metabolic footprinting and systems biology: the medium is the message. *Nat Rev Microbiol* **3**, 557–565, <https://doi.org/10.1038/nrmicro1177> (2005).
- Leon, Z., Garcia-Canaveras, J. C., Donato, M. T. & Lahoz, A. Mammalian cell metabolomics: experimental design and sample preparation. *Electrophoresis* **34**, 2762–2775, <https://doi.org/10.1002/elps.201200605> (2013).
- Nunes de Paiva, M. J., Menezes, H. C. & de Lourdes Cardeal, Z. Sampling and analysis of metabolomes in biological fluids. *Analyst* **139**, 3683–3694, <https://doi.org/10.1039/c4an00583j> (2014).
- Silva, C. L., Perestrello, R., Silva, P., Tomas, H. & Camara, J. S. Volatile metabolomic signature of human breast cancer cell lines. *Sci Rep* **7**, 43969, <https://doi.org/10.1038/srep43969> (2017).
- Poli, D. *et al.* Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatization SPME-GC/MS. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **878**, 2643–2651, <https://doi.org/10.1016/j.jchromb.2010.01.022> (2010).
- Li, J. *et al.* Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. *Clinica chimica acta; international journal of clinical chemistry* **436**, 59–67, <https://doi.org/10.1016/j.cca.2014.04.030> (2014).
- Leunis, N. *et al.* Application of an electronic nose in the diagnosis of head and neck cancer. *The Laryngoscope* **124**, 1377–1381, <https://doi.org/10.1002/lary.24463> (2014).

34. Kumar, S. *et al.* Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer. *Analytical chemistry* **85**, 6121–6128, <https://doi.org/10.1021/ac4010309> (2013).
35. Altomare, D. F. *et al.* Exhaled volatile organic compounds identify patients with colorectal cancer. *The British journal of surgery* **100**, 144–150, <https://doi.org/10.1002/bjs.8942> (2013).
36. Abaffy, T., Moller, M. G., Riemer, D. D., Milikowski, C. & DeFazio, R. A. Comparative analysis of volatile metabolomics signals from melanoma and benign skin: a pilot study. *Metabolomics: Official journal of the Metabolomic Society* **9**, 998–1008, <https://doi.org/10.1007/s11306-013-0523-z> (2013).
37. Mochalski, P. *et al.* Release and uptake of volatile organic compounds by human hepatocellular carcinoma cells (HepG2) *in vitro*. *Cancer cell international* **13**, 72, <https://doi.org/10.1186/1475-2867-13-72> (2013).
38. Khalid, T. *et al.* Urinary Volatile Organic Compounds for the Detection of Prostate Cancer. *PLoS one* **10**, e0143283, <https://doi.org/10.1371/journal.pone.0143283> (2015).
39. Steve Smith, P. W., Redding, J., Ratcliffe, N. M. & Chris S. J. Probert. Application of Similarity Coefficients to Predict Disease Using Volatile Organic Compounds. *IEEE SENSORS JOURNAL* **10**, 92–96, <https://doi.org/10.1109/JSEN.2009.2035771> (2010).
40. Huang, S. *et al.* Acidic extracellular pH promotes prostate cancer bone metastasis by enhancing PC-3 stem cell characteristics, cell invasiveness and VEGF-induced vasculogenesis of BM-EPCs. *Oncol Rep* **36**, 2025–2032, <https://doi.org/10.3892/or.2016.4997> (2016).
41. Marcias, G. *et al.* Identification of novel truncated androgen receptor (AR) mutants including unreported pre-mRNA splicing variants in the 22Rv1 hormone-refractory prostate cancer (PCa) cell line. *Human mutation* **31**, 74–80, <https://doi.org/10.1002/humu.21138> (2010).
42. Wishart, D. S. *et al.* HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic acids research* **41**, D801–807, <https://doi.org/10.1093/nar/gks1065> (2013).
43. Wishart, D. S. *et al.* HMDB: the Human Metabolome Database. *Nucleic acids research* **35**, D521–526, <https://doi.org/10.1093/nar/gkl923> (2007).
44. Wishart, D. S. *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic acids research* **37**, D603–610, <https://doi.org/10.1093/nar/gkn810> (2009).
45. Zimmermann, D., Hartmann, M., Moyer, M. P., Nolte, J. & Baumbach, J. I. Determination of volatile products of human colon cell line metabolism by GC/MS analysis. *Metabolomics: Official journal of the Metabolomic Society* **3**, 13–17, <https://doi.org/10.1007/s11306-006-0038-y> (2007).
46. Sponring, A. *et al.* Analysis of volatile organic compounds (VOCs) in the headspace of NCI-H1666 lung cancer cells. *Cancer biomarkers: section A of Disease markers* **7**, 153–161, <https://doi.org/10.3233/CBM-2010-0182> (2010).
47. Wojciech, F. *et al.* A compendium of volatile organic compounds (VOCs) released by human cell lines. *Current medicinal chemistry* (2016).
48. Erhart, S. *et al.* 3-Heptanone as a potential new marker for valproic acid therapy. *J Breath Res* **3**, 016004, <https://doi.org/10.1088/1752-7155/3/1/016004> (2009).
49. Madhu, B., Shaw, G. L., Warren, A. Y., Neal, D. E. & Griffiths, J. R. Response of Degarelix treatment in human prostate cancer monitored by HR-MAS 1H NMR spectroscopy. *Metabolomics: Official journal of the Metabolomic Society* **12**, 120, <https://doi.org/10.1007/s11306-016-1055-0> (2016).
50. Hakim, M. *et al.* Volatile organic compounds of lung cancer and possible biochemical pathways. *Chemical reviews* **112**, 5949–5966, <https://doi.org/10.1021/cr300174a> (2012).
51. Filipiak, W. *et al.* Release of volatile organic compounds (VOCs) from the lung cancer cell line CALU-1 *in vitro*. *Cancer cell international* **8**, 17, <https://doi.org/10.1186/1475-2867-8-17> (2008).
52. Davies, M. P. *et al.* Unique volatolomic signatures of TP53 and KRAS in lung cells. *British journal of cancer* **111**, 1213–1221, <https://doi.org/10.1038/bjc.2014.411> (2014).
53. Muzio, G., Maggiora, M., Paiuzzi, E., Oraldi, M. & Canuto, R. A. Aldehyde dehydrogenases and cell proliferation. *Free Radic Biol Med* **52**, 735–746, <https://doi.org/10.1016/j.freeradbiomed.2011.11.033> (2012).
54. Yan, J., De Melo, J., Cutz, J. C., Aziz, T. & Tang, D. Aldehyde dehydrogenase 3A1 associates with prostate tumorigenesis. *British journal of cancer* **110**, 2593–2603, <https://doi.org/10.1038/bjc.2014.201> (2014).
55. Peng, G. *et al.* Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. *British journal of cancer* **103**, 542–551, <https://doi.org/10.1038/sj.bjc.6605810> (2010).
56. Taware, R. *et al.* Investigation of urinary volatolomic alterations in head and neck cancer: a non-invasive approach towards diagnosis and prognosis. *Metabolomics: Official journal of the Metabolomic Society* **13**, 111, <https://doi.org/10.1007/s11306-017-1251-6> (2017).
57. Wallqvist, A. *et al.* Mining the NCI screening database: explorations of agents involved in cell cycle regulation. *Progress in cell cycle research* **5**, 173–179 (2003).
58. Haick, H., Broza, Y. Y., Mochalski, P., Ruzsanyi, V. & Amann, A. Assessment, origin, and implementation of breath volatile cancer markers. *Chemical Society reviews* **43**, 1423–1449, <https://doi.org/10.1039/c3cs60329f> (2014).
59. Tang, H. *et al.* Determination of volatile organic compounds exhaled by cell lines derived from hematological malignancies. *Biosci Rep* **37**, <https://doi.org/10.1042/BSR20170106> (2017).
60. Zhang, Y. *et al.* Identification of volatile biomarkers of gastric cancer cells and ultrasensitive electrochemical detection based on sensing interface of Au-Ag alloy coated MWCNTs. *Theranostics* **4**, 154–162, <https://doi.org/10.7150/thno.7560> (2014).
61. Wu, H. *et al.* GC/MS-based metabolomic approach to validate the role of urinary sarcosine and target biomarkers for human prostate cancer by microwave-assisted derivatization. *Anal Bioanal Chem* **401**, 635–646, <https://doi.org/10.1007/s00216-011-5098-9> (2011).
62. Zang, X. *et al.* Feasibility of detecting prostate cancer by ultraperformance liquid chromatography-mass spectrometry serum metabolomics. *Journal of proteome research* **13**, 3444–3454, <https://doi.org/10.1021/pr500409q> (2014).
63. Kwak, J. *et al.* Volatile biomarkers from human melanoma cells. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* **931**, 90–96, <https://doi.org/10.1016/j.jchromb.2013.05.007> (2013).
64. Huang, Z. *et al.* Effects of culture media on metabolic profiling of the human gastric cancer cell line SGC7901. *Mol Biosyst* **11**, 1832–1840, <https://doi.org/10.1039/c5mb00019j> (2015).
65. Monteiro, M. *et al.* Analysis of volatile human urinary metabolome by solid-phase microextraction in combination with gas chromatography-mass spectrometry for biomarker discovery: application in a pilot study to discriminate patients with renal cell carcinoma. *Eur J Cancer* **50**, 1993–2002, <https://doi.org/10.1016/j.ejca.2014.04.011> (2014).
66. Mastrangelo, A., Ferrarini, A., Rey-Stolle, F., Garcia, A. & Barbas, C. From sample treatment to biomarker discovery: A tutorial for untargeted metabolomics based on GC-(EI)-Q-MS. *Analytica chimica acta* **900**, 21–35, <https://doi.org/10.1016/j.aca.2015.10.001> (2015).
67. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395, <https://doi.org/10.1186/1471-2105-11-395> (2010).
68. Jentzmik, F. *et al.* Sarcosine in urine after digital rectal examination fails as a marker in prostate cancer detection and identification of aggressive tumours. *Eur Urol* **58**, 12–18, <https://doi.org/10.1016/j.eururo.2010.01.035> (2010). discussion 20–11.
69. Berben, L., Sereika, S. M. & Engberg, S. Effect size estimation: methods and examples. *International journal of nursing studies* **49**, 1039–1047, <https://doi.org/10.1016/j.ijnurstu.2012.01.015> (2012).

## Acknowledgements

This work received financial support from the European Union (FEDER funds POCI/01/0145/FEDER/07728) and National funds (FCT/MEC, Fundação para a Ciência e a Tecnologia and Ministério da Educação e Ciência) under the Partnership Agreement PT2020 UID/MULTI/04378/2013. A.R.L. and A.M.A thank Fundação para a Ciência e Tecnologia for their grants, respectively SFRH/BD/123012/2016 and SFRH/BD/107708/2015.

## Author Contributions

A.R.L. carried out the experiments, analyzed data, and wrote the manuscript. A.M.A. supported cell culture and data analysis. J.P. helped with the statistical analysis of the data. C.J. and R.H. kindly provided the cell lines used in the study and gave conceptual advice. P.G.P., M.C. and M.L.B. designed the experiments, supervised the project, and revised the manuscript. All authors critically commented on and approved the final submitted version of the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-23847-9>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018