Volume 144 | Number 14 | 21 July 2019 | Pages 4125–4448

Analyst

rsc.li/analyst

ISSN 0003-2654

ROYAL SOCIETY OF CHEMISTRY | Celebrating IYPT 2019

PAPER
José S. Câmara *et al.*
Volatomic pattern of breast cancer and cancer-free tissues as a powerful strategy to identify potential biomarkers

Analyst

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2019, **144**, 4153

Volatomic pattern of breast cancer and cancer-free tissues as a powerful strategy to identify potential biomarkers†

Catarina Silva, [a] Rosa Perestrelo, [a] Pedro Silva, [a] Filipa Capelinha,[b] Helena Tomás [a,c] and José S. Câmara [a,c]*

Received 7th February 2019,
Accepted 11th May 2019

DOI: 10.1039/c9an00263d

rsc.li/analyst

Breast cancer (BC), ranked as the fifth amongst all cancers, remains at the top of women's cancers worldwide followed by colorectal, lung, cervix, and stomach cancers. The main handicap of most of the screening/diagnostic methods is based on their low sensitivity and specificity and the invasive behavior of most sampling procedures. The aim of this study was to establish the volatomic pattern of BC and cancer-free (CF) tissues (n = 30) from the same patients, as a powerful tool to identify a set of volatile organic metabolite (VOM) potential BC biomarkers which might be used together or complement with the traditional BC diagnostics strategies, through the integration of chromatographic data, obtained by solid-phase microextraction followed by gas chromatography-mass spectrometry (SPME/GC-qMS), with chemometric tools. A total of four metabolites: limonene, decanoic acid, acetic acid and furfural presented the highest contribution towards discrimination of BC and CF tissues (VIP > 1, p < 0.05). The discrimination efficiency and accuracy of BC tissue metabolites was ascertained by ROC curve analysis that allowed the identification of some metabolites with high sensitivity and specificity. The results obtained with this approach suggest the possibility of identifying endogenous metabolites as a platform to find potential BC biomarkers and pave the way to investigate the related metabolomic pathways in order to improve BC diagnostic tools. Moreover, deeper investigations could unravel novel mechanistic insights into the disease pathophysiology.

Introduction

Breast cancer (BC) is ranked as the fifth amongst all cancers remaining at the top of women's cancers worldwide followed by colorectal, lung, cervix, and stomach cancers according to the GLOBOCAN series of the International Agency for Research on Cancer (IARC), contributing to more than 11.6% of all cancers.[1] Although BC is a multifactorial disease, with highly variable clinical behavior and response to therapy, it can be curable in early stages. Furthermore, there is still the need for the development of new methodologies to aid or monitor the disease together with the current diagnostic tools, namely mammography, ultrasound or tumor markers. Moreover, before BC treatment, a complex and time-consuming analysis is required that uses many different assays, such as the determination of histological type and grading and the evaluation of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER-2), among others.[2] The main handicap of most of these screening/diagnostic methods is their low sensitivity and specificity and the invasive procedure required to obtain the samples.[3] Taking into account these aspects, research is being directed towards the use of new tools that can support the clinicians in BC treatment and follow-up.[4] In this sense, in recent years, metabolomic studies have emerged as a powerful tool to investigate the changes and/or metabolic responses of living systems to stimuli or genetic modifications.[5] The metabolome profile represents the unbiased quantitative and qualitative analyses of the complete set of metabolites present in cells, body fluids and/or tissues.[6] To date, beyond the most used biological specimens (*e.g.* urine, saliva, blood), BC tissues have been used in metabolomics with the aim of discriminating cancer from normal tissues suggesting that metabolomic profiles differ within molecular subtypes of BC.[7,8] The metabolome

[a]*CQM, Centro de Química da Madeira, Universidade da Madeira, Campus da Penteada, 9020-105 Funchal, Portugal*
[b]*SESARAM, EPE. Hospital Dr. Nélio Mendonça, Serviço de Anatomia Patológica, Avenida Luís de Camões, nº 57–9004-514 Funchal, Portugal.*
E-mail: jsc@staff.uma.pt
[c]*Faculdade de Ciências Exatas e da Engenharia, Universidade da Madeira, Campus da Penteada, 9020-105 Funchal, Portugal*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c9an00263d

coverage in BC tissues can be maximized by combining different technologies for metabolic profiling, namely gas chromatography mass-spectrometry (GC-qMS), and the results can be used to classify BC helping to identify new prognostic and predictive markers and to discover new targets for future therapeutic interventions.[9] Among them, the study of volatile organic metabolites (VOMs) present in biological samples, namely in saliva, urine, breath and tissues can be useful for cancer diagnosis, in particular for BC. The most common procedure used in the extraction of volatile compounds in biological samples is solid-phase microextraction (SPME), normally in the headspace mode (HS-SPME) being used in several biological matrices.[10–13] GC-qMS was used to screen salivary volatiles for putative BC as an exploratory study involving geographically distant populations,[14] also to establish the metabolomic signature of human BC cell lines[11] and to discriminate different types of cancer based on urinary volatomic biosignatures.[13] In the first study, up to 120 VOMs from distinct chemical families, with significant variations among the groups, were identified,[14] whereas Silva *et al.*[11] and Porto-Figueira *et al.*[13] identified 60 and 130 VOMs in BC cell lines and urine, respectively. On the other hand, Budczies *et al.*[15,16] used a gas chromatography time of flight mass spectrometry (GC-TOFMS) framework to evaluate the glutamate enrichment as a new diagnostic opportunity in BC and to accomplish comparative metabolomics of estrogen receptor positive (ER^+) and estrogen receptor negative (ER^-) samples. Budczies *et al.*[16] identified 19 VOMs and the GC-TOFMS based analysis of metabolites present in BC tissues revealed significant differences in central metabolism in the more aggressive ER^- compared to the ER^+ type. The detected changes included the metabolism of glutamine with a decrease in the concentration of glutamine and an increase in the concentration of glutamate and 2-hydroxyglutaric acid.[16] In turn, Dougan *et al.*[8] used GC-MS to evaluate the detectability, reliability, and distribution of metabolites measured in pre-diagnostic plasma samples in a pilot study of women listed in the Northern California site of the BC Family Registry. In this study, 661 VOMs were detected, 338 (51%) of them were found in all samples, and 490 (74%) in more than 80% of samples. The aim of this study was to establish the volatomic pattern of BC tissue and CF tissue samples collected after surgery, from the same patient (to minimize the interference of epigenetic and external factors), in order to find a set of volatile metabolites to be used as potential BC biomarkers, using HS-SPME/GC-qMS combined with multivariate statistical tools. This high-throughput strategy might have the potential to be applied in a clinical environment as a diagnostic approach or as a complementary way with the current diagnostic methods to improve the diagnostic decision.

Experimental

Materials and reagents

Sodium chloride (NaCl), hydrochloric acid (HCl) and 4-methyl-2-pentanol were supplied by Panreac (Barcelona, Spain) and Sigma Aldrich (St Louis, MO, USA), respectively. The C_8–C_{20} alkane solution (concentration of 40 mg L^{-1} in *n*-hexane) was purchased from Fluka (Buchs, Switzerland). The digital stirring plate (Cimarec™) was supplied by Thermo Scientific (Waltham, MA, USA) while the SPME holder for manual sampling and 75 μm Carboxen/polydimethylsiloxane fiber (CAR/PDMS) were purchased from Supelco (Bellefonte, PA, USA).

Subjects and tissue collection

To investigate the BC tissue metabolomic profile, 30 samples from patients with breast cancer (BC, n = 30, age range 44–85, average 67), and 30 samples from cancer-free tissues (CF, n = 30, age range 44–85, average 67) without malignant infiltration were resected from each patient. The resected samples were divided into the active carcinoma and cancer-free tissues outside the tumor margin and were immediately frozen in liquid nitrogen, in a total set of 60 samples. The tissues were stored at −80 °C until extraction. These samples were obtained at the Pathological Anatomy Unit of Hospital Dr Nélio Mendonça (Funchal, Portugal) according to Table 1. All experiments were performed in accordance with the standard Guidelines from Declaration of Helsinki and approved by the institutional ethics committee of Hospital Dr Nélio Mendonça and University of Madeira. All the participants of this study were informed about the investigation and a signed informed consent was obtained from all human participants prior to sample collection. Using the TNM (tumor, node, and metastasis) staging approach, the examined BC cases included five of stage IA, ten of stage IIA, one of stage IIIA B, seven of stage IIB, five of stage IIIB and two of stage IIIC.

Extraction of metabolites from breast tissues

The HS-SPME extraction conditions were based on a developed method previously established in our laboratory.[10,12] Briefly, tissue samples were thawed and then portions of 100 mg were weighed into 20 mL vials together with 17% NaCl (w/v), 1000 μL of ultrapure water and 100 μL of the internal standard (IS, 4-methyl-2-pentanol, 1.6 mg L^{-1}). The pH was adjusted to 2 with small amounts of 5 M HCl. Then, the vial was capped with a Teflon (PTFE) septum using a screw cap and the SPME fiber was introduced and exposed into the headspace over 75 min at 50 °C at 800 rpm (0.5 mm × 0.1 mm bar). After this period, the fiber was removed from the vial and inserted into the GC injection port and the VOMs extracted were desorbed

Table 1 List of collected tissue samples from breast cancer patients

Samples	BC tissue	Cancer-free tissue
Number	30	30
Age (range, median)	(44–85, 65)	(44–85, 65)
Histological grade (number of samples)	IA (5) IIA (10) IIIA (1) IIB (7) IIIB (5) IIIC (2)	Not applicable

for 10 min at 250 °C. Each sample was analyzed in duplicate and blanks were performed before each analysis.

Gas chromatography quadrupole mass-spectrometry (GC-qMS) conditions

After the extraction procedure, the SPME fiber with the analytes was inserted into the injection port of an Agilent Technologies 6890N Network gas chromatograph system (Palo Alto, CA, USA) where the VOMs were desorbed at 250 °C for 10 min. The gas chromatograph was equipped with a 60 m × 0.25 mm I.D. × 0.25 μm film thickness, BP-20 (SGE, Dortmund, Germany) fused silica capillary column and interfaced with an Agilent 5975 quadrupole inert mass selective detector. The following oven temperature profile was set: (a) 5 min at 45 °C; (b) temperature was increased to 150 °C, at a rate of 2 °C min^{-1} (hold for 10 min); (c) 150 °C for 10 min; (d) temperature was increased to 220 °C, at a rate of 7 °C min^{-1}; and (e) 220 °C for 10 min for a total GC run time of 87.5 min. The column flow was constant at 1.3 mL min^{-1} using helium (He, N60, Air Liquide, Portugal) as a carrier gas. The injection port was operated in the splitless mode and held at 250 °C. For the 5975 MS system, the operating temperatures of the transfer line, quadrupole and ionization source were 270, 150 and 230 °C, respectively, while electron impact mass spectra were recorded at 70 eV ionization voltage and the ionization current was 10 μA. Data acquisition was performed in the scan mode (30–200 *m*/*z*). The electron multiplier was set to the autotune procedure. Metabolite identification was accomplished by manual interpretation through single ion monitoring (SIM) of spectra and matching against the Agilent MS ChemStation Software, equipped with a NIST05 mass spectral library with a similarity threshold higher than 80% and comparison with commercially available standard samples when available. A series of C_8–C_{20} *n*-alkanes were analyzed using the same extraction procedure to establish the Kovats indices (KI), and to confirm the identity of the VOMs by comparison with the literature. The analyses were performed in duplicate and the results are expressed by means ± standard deviation.

Statistical analysis

Statistical analysis was performed using the web server Metaboanalyst 4.0.[17] The multivariate statistical analysis, namely the principal component analysis (PCA), the partial least squares-discriminant analysis (PLS-DA) and the orthogonal projections to latent structures discriminant analysis (OPLS-DA) were applied on the tissue metabolomic profile dataset to provide insights into the groups under study. The metabolites with VIP scores higher than 1.0 selected by the PLS-DA analysis were used for the pathway analysis. Furthermore, hierarchical cluster analysis by K-means of the two groups in study was carried out and Pearson's correlation was used to build the heat map with the aim of identifying BC clustering patterns. The receiver operating characteristic (ROC) curves were attained to verify which metabolites had the highest sensitivity/specificity for a potential BC diagnosis. Finally, the selected metabolites were used for the metabolic pathway analysis to identify the most relevant metabolic pathways involved in BC.

Results and discussion

Tissue metabolomic pattern based on GC-qMS

A total of twenty-nine VOMs were identified in BC tissue and CF tissue samples which were classified in several chemical families, namely phenols, benzene derivatives, carbonyl compounds, acids, alcohols and furanic compounds as shown in ESI Fig. S1.† Data were processed using software (NIST, 2005; Mass Spectral Search Program V.2.0d) which provides quality matching using advanced spectral matching algorithms background subtraction and KI comparison. The KI were determined through the injection of a C_8–C_{20} alkane solution in order to confirm the identity of the VOMs by comparison with the literature (*e.g.* pherobase). It can be observed in Table 2 that the KI of the identified VOMs, when compared to the ones found in the literature for a similar column (BP-20), are closer to one another, ensuring that a good identification was achieved.

The highest contribution for the volatile profile was from phenols in the BC group, whereas for the CF group it was from benzene derivatives and acids. The main VOMs identified for these families were phenol, toluene and acetic acid, respectively. These metabolites were already identified in several reports in the literature in biological matrices such as urine,[10,12] cancer cell lines,[11,18] saliva,[14] and exhaled breath.[18,19] Several studies reported the potential of VOMs from different biological matrices on differentiation between cancer patients and healthy subjects. In addition, different endogenous metabolites were identified as potential cancer biomarkers which might be used in cancer diagnosis. These metabolites can be analysed through different analytical platforms including GC-MS, LC-MSMS and NMR. Table 2 shows the identification of VOMs, as well as the minimum and maximum relative GC peak areas for each metabolite and group. It can be observed that most metabolites were identified in all samples with a frequency of occurrence (FO) higher than 90%, where the maximum relative area was obtained for *p-tert*-butyl-phenol in the CF group and phenol for the BC group. For the CF group, the major VOMs identified include acetic acid, phenol and *p-tert*-butyl-phenol.

Some of these have been already reported in the literature, namely acetic acid was found as a discriminant when associated with controls (CTL) in a study conducted by Ahmed *et al.*[20] that investigated the possibility of faecal VOMs as potential diagnostic biomarkers for inflammatory bowel disease. Mochalski *et al.*[21] investigated the emission of volatile compounds from gastric cancer tissues and non-cancerous tissues with the aim of identifying characteristic chemical patterns associated with gastric cancer. Furthermore, Silva *et al.*[10,12] also reported phenols as the major chemical family identified in urine from the oncologic group. Furthermore, Raman *et al.*[22] studied faecal VOMs in obese humans and

Table 2 Identified metabolites of tissue samples from BC patients and cancer-free (CF) through GC-qMS, minimum (Min) and maximum (Max), variation of relative peak areas (relative to internal standard, RSD < 10%) regarding the BC group and their frequency of occurrence (in %). Identification mode and the Kovats index for each identified VOM and the literature values for a similar GC column

							Relative GC peak areas					
							CF		BC			
Peak no.	IEC[a]	RT[b] (min)	ID[c]	KI_{cal} [d]	KI_{lit} [e]	VOM[f]	Min	Max	Min	Max	Variation	Frequency of occurrence (FO)/%
A1	41, 69	9.77	MS	990	983	Methyl isobutyrate	0.03	1.81	0.02	1.66	↓	71
A2	91	11.24	Std, MS	1029	1042	Toluene	0.03	7.62	0.01	17.17	↑	100
A3	44, 56	13.12	Std, MS	1073	1075	Hexanal	0.03	6.82	0.04	2.09	↓	99
A7	41, 70	18.48	Std, MS	1172	1168	Heptanal	0.04	0.17	0.05	0.19	↑	13
A8	68, 93	18.73	St, MS	1176	1198	Limonene	0.01	4.46	0.01	1.03	↓	92
A10	43, 33	19.84	St, MS	1187	—	2-Methyl-1-propanol	0.01	1.44	0.01	0.51	↓	79
A11	81	20.69	MS	1197	1229	2-Pentyl-furan	0.01	0.07	0.01	0.13	↑	77
A12	105, 120	21.39	MS	1209	—	Trimethylbenzene	0.01	5.41	0.01	1.70	↓	64
A13	119, 134	23.18	St, MS	1237	1234	*o*-Cymene	0.01	0.84	0.01	0.33	↓	24
A14	42, 55	24.36	MS	1254	1255	1-Pentanol	0.01	0.03	0.01	0.23	↑	50
A15	43, 56	24.90	St, MS	1262	1280	Octanal	0.01	0.24	0.01	9.19	↑	34
A16	56	30.20	St, MS	1338	1360	1-Hexanol	0.01	0.05	0.01	1.17	↑	27
A17	41, 57	31.89	St, MS	1362	1385	Nonanal	0.02	0.48	0.17	2.36	↑	52
A18	117, 132	34.45	MS	1396	—	*p*-Cymenene	0.01	0.76	0.01	0.15	↓	100
A19	43, 60	36.42	St, MS	1428	1450	Acetic acid	0.62	42.75	0.10	12.28	↓	100
A20	96	36.61	St, MS	1431	1455	Furfural	0.01	5.87	0.04	18.03	↑	100
A21	57	38.53	St, MS	1461	1487	2-Ethyl-1-hexanol	0.01	0.50	0.01	3.20	↑	2
A22	43, 57	38.77	St, MS	1465	1484	Decanal	n.d.[g]	n.d.	0.14	0.15	—	72
A23	77, 106	39.81	St, MS	1480	1495	Benzaldehyde	0.01	0.36	0.01	0.33	↓	54
A24	41, 56	42.63	MS	1525	1553	1-Octanol	0.01	0.17	0.01	1.57	↑	74
A25	77, 105	47.36	MS	1602	1607	Acetophenone	0.01	0.19	0.01	0.42	↑	79
A26	60, 73	58.70	St, MS	1802	1829	Hexanoic acid	0.01	1.25	0.02	5.77	↑	100
A27	94	69.54	St, MS	1968	1965	Phenol	0.51	23.32	0.45	44.08	↑	39
A28	60, 73	72.22	St, MS	2004	2013	Octanoic acid	0.01	0.25	0.01	6.51	↑	76
A29	107	73.00	MS	2009	2017	4-Methyl-phenol	0.01	1.08	0.01	3.72	↑	45
A30	60, 73	75.65	St, MS	2027	—	Nonanoic acid	0.01	0.18	0.01	0.64	↑	29
A31	155, 170	75.71	MS	2027	—	1,6,7-Trimethyl-naphthalene	0.02	0.41	0.01	0.37	↓	39
A32	60, 73	78.15	St, MS	2043	—	Decanoic acid	0.01	0.11	0.02	0.52	↑	100
A33	135, 150	78.49	St, MS	2045	—	*p-tert*-Butyl-phenol	0.02	53.76	0.03	20.71	↓	100

[a] IEC – ion extraction chromatogram. [b] Retention time (min). [c] ID – metabolite identification using a standard compound (St) or mass spectra of the NIST library search (MS). [d] Kovats index relative to *n*-alkanes (C_8–C_{20}) on a BP-20 capillary column. [e] Kovats index relative to that reported in the literature for an equivalent capillary column. [f] VOM – volatile organic metabolite. [g] n.d. – not detected.

identified also acetic acid and phenol as major metabolites. On the other hand, P. Porto-Figueira *et al.*[13] also identified *p*-xylene, *o*-xylene, acetic acid, phenol and *p-tert*-butyl-phenol in urine from cancer patients, with some of them presenting higher values in cancer patients. Regarding the relative area values obtained for the BC group, most VOMs were down-regulated with the exception of acetic acid, toluene, octanal and nonanal.

Multivariate statistical analysis of tissue metabolomic profile

The statistical analysis was performed using the Metaboanalyst 4.0[17] web server as described in the Experimental section. Only the VOMs with FO higher than 90% were considered for the statistical analysis, in a total of 8 VOMs (toluene, hexanal, limonene, *p*-cymenene, acetic acid, furfural, hexanoic acid and decanoic acid). Initially, data were transformed by log transformation and mean centering approaches, before being subjected to multivariate statistical analysis. Partial least squares-discriminant analysis (PLS-DA) was used as a supervised clustering method to verify the existence of an altered metabolite pattern. Additionally, this type of statistical analysis takes into account the variance/covariance between samples of groups where the samples are classified into different groups (ESI Fig. S2A†). Then, the top four metabolites (limonene, decanoic acid, acetic acid and furfural) with the highest contribution for group discrimination were selected with variable importance in projection (VIP > 1) (ESI Fig. S2B†). Furthermore, Fig. 1 shows the comparison of three illustrative VOMs with a higher VIP score from BC tissue and CF tissue. It can be observed that for a majority of metabolites their level is higher in the CF group rather than in the BC group. Moreover, orthogonal partial least squares-discriminant (OPLS-DA) analysis was applied on the tissue metabolomic profile dataset to maximize the separation of BC and CF groups. Significant group separation was observed in the OPLS-DA score plot between BC patients and CF group indicating intrinsic metabolic alterations in each group (Fig. 2A). To attest the robustness of the model, a random permutation test with 1000 permutations was performed with OPLS-DA (Fig. 2B). The permutation test yielded R^2 (represents goodness of fit) as 0.717 and Q^2 (represents predictive ability) as 0.691 indicating that the model is not over fitted and has a relative good predictive ability to distinguish between study groups.

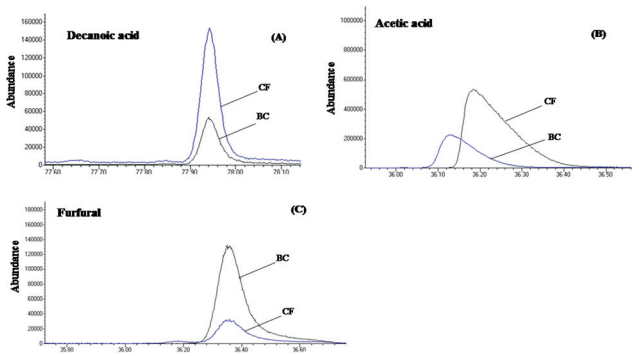

Fig. 1 Comparison between BC and CF groups of three illustrative VOMs selected by VIP scores.

Although in the literature there are no reports regarding the volatile profile of human tissues, many of these metabolites were already identified in previous reports, namely in lung,[23,24] breast, ovarian,[3,25] bladder,[26] and gastric cancers.[27] The OPLS-DA uses class information allowing us to show which variables are responsible for class discrimination using the predictive information of the first component. The main advantage of OPLS-DA when compared to PLS-DA is that a single component is used as a predictor for the class, while the other components describe the variation orthogonal to the first predictive component.[28]

To further evaluate the predictive value of the metabolites to discriminate between BC patients and CF, a receiver operating characteristic curve (ROC) analysis was generated using the top four metabolites identified by VIP values (Fig. 3A and B). This type of analysis is used for the classification of true positives and false positives and the predictive ability is measured using the area under the curve (AUC).[29,30] According to Xia *et al.*,[31] an AUC between 0.9–1.0 is excellent and between 0.8–0.9 is good. Based on this classification, the results obtained were, thus, very good (AUC = 0.966).

ROC curves were generated by Monte-Carlo cross validation (MCCV) using balanced sub-sampling. In each MCCV, two thirds (2/3) of the samples were used to evaluate the feature importance. The top 3 important features were then used to build classification models which were then validated on 1/3 of the samples that were left out. This procedure was repeated multiple times to calculate the performance and confidence interval of each model. The AUC can be interpreted as the probability that a randomly selected diseased subject is ranked as more probable to be diseased than a randomly selected healthy subject.[31] A greater AUC value indicates the effectiveness to separate the CF group from the cancer group (BC).

Additionally, a 10-fold cross validation was used to generate a logistic regression model and the performance was calculated according to the equation:

$$\mathrm{logit}(P) = \log(P/(1-P)) = -0.077 + 0.966\ \text{limonene} - 0.752\ \text{acetic acid} - 0.076\ \text{furfural} - 0.133\ \text{decanoic acid},$$

where P is $\Pr(y = 1|x)$. The threshold (or cut-off) for the predicted P was 0.61 (ESI Tables S1 and S2†). Fig. 4A and B show the results obtained for the predicted probabilities using the OPLS-DA model and the average of the predictive accuracy for the same model, where it can be observed that the model allowed a good classification of samples (>90%). Moreover, 20 samples without known labels were processed together with the ones with known labels in order to obtain the probability of class labels as presented in ESI Table S3.† Most of the cases were classified in

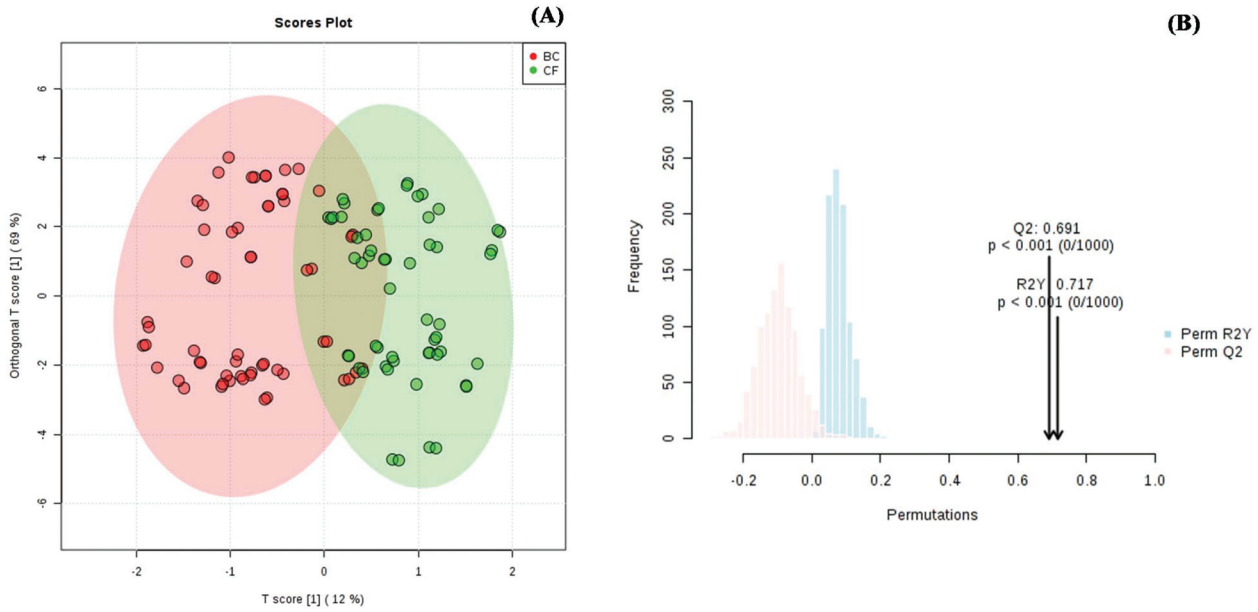

Fig. 2 Loading score plots of orthogonal projection to latent structure discriminant analysis (OPLS-DA) and (B) model validation by the permutation test based on 1000 permutations of VOMs obtained by GC-qMS of tissue samples from the 2 groups under study.

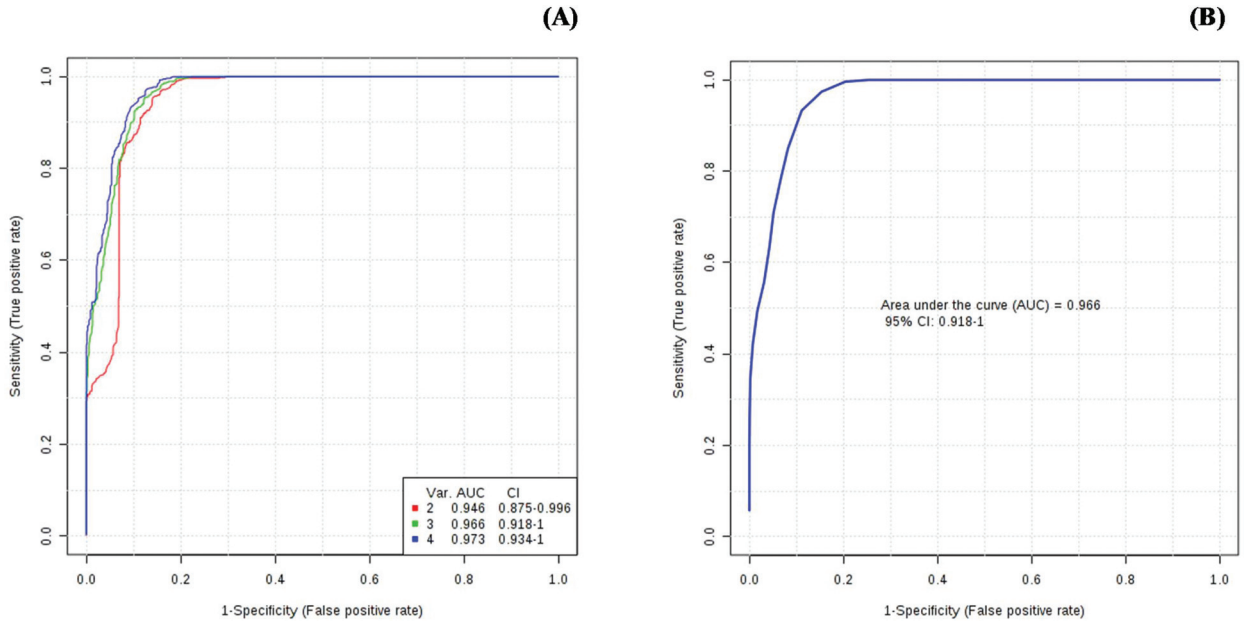

Fig. 3 Receiver operating characteristic (ROC) curves for the predictive model. (A) A combination metabolites model calculated from the logistic regression analysis using the 4 metabolites selected by the VIP (>1.0) values, (B) ROC curve for the top 4 metabolites (limonene, decanoic acid, acetic acid and furfural) with the highest ability to discriminate BC patients against the CF.

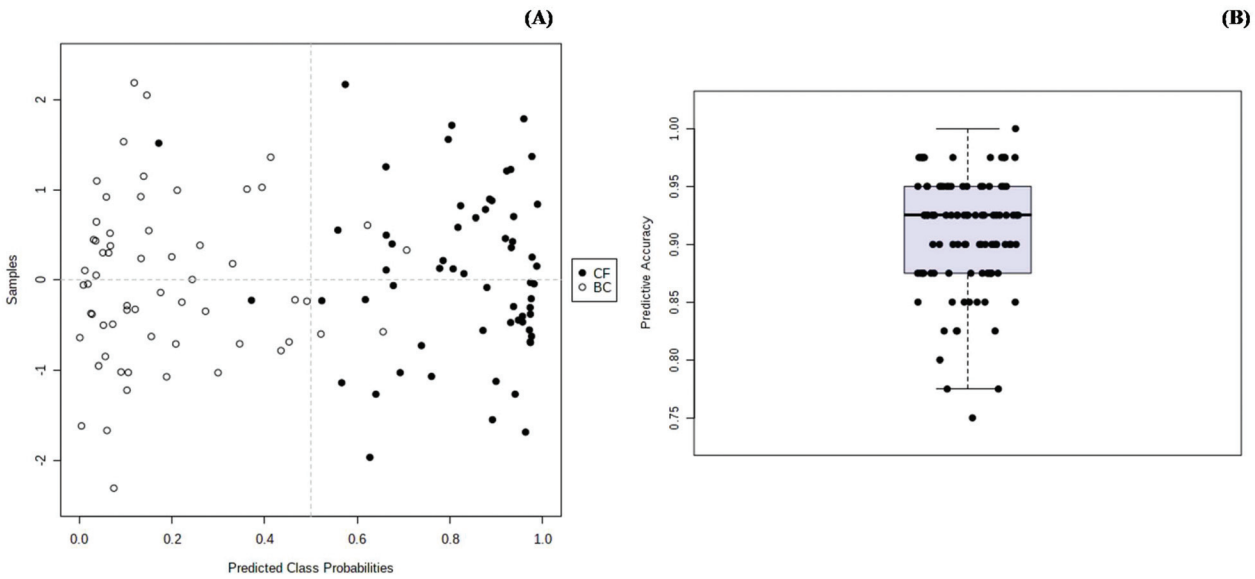

Fig. 4 (A) Plot of the predicted class probabilities for all samples using the OPLS-DA biomarker model based on AUC and (B) box plot of the predictive accuracy (with an average of 0.908) of the biomarker model based on 100 cross validations.

their respective groups with the exception of O10, O5 and O18 with a probability score ranging from 0.982 to 0.595.

Also, the heat map was constructed with selected VOMs by VIP > 1, using Pearson's correlation, providing intuitive visualization of the data set and the correlations between samples and VOMs (Fig. 5).

Finally, the metabolic pathway analysis was performed using the metabolites with VIP values higher than 1 in order to explore which pathways changed in BC. Fig. 6A shows the most important pathways with the highest impact on BC. As shown, sulfur and pyruvate metabolisms were the most important metabolic pathways explaining the BC carcinogenesis. In Fig. 6B we can observe the most important metabolomics routes producing the most significant volatile metabolites in BC (indicated by KEGG accession number).

It can be observed that the pathway with the highest impact was the pyruvate and sulfur metabolism due to acetic acid. This metabolite according to the human metabolome database

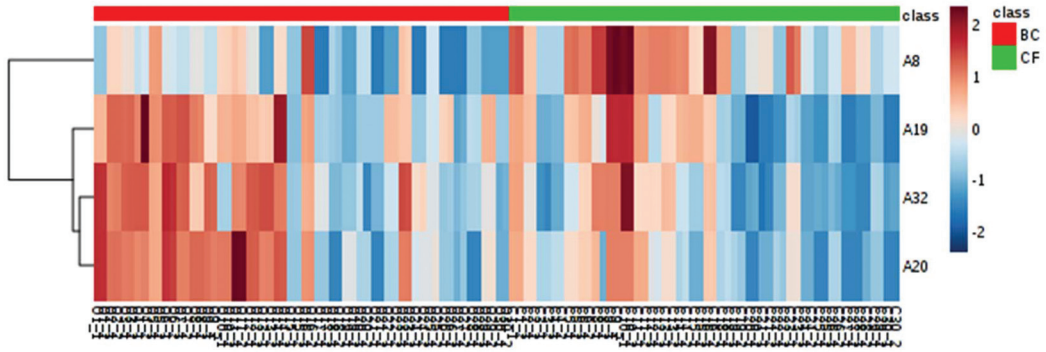

Fig. 5 Heat map visualization and hierarchical clustering analysis using the four metabolites with significance ($p < 0.05$) by Pearson's distance analysis (A8 – limonene; A19 – acetic acid; A32 – decnoic acid: A20 – furfural).

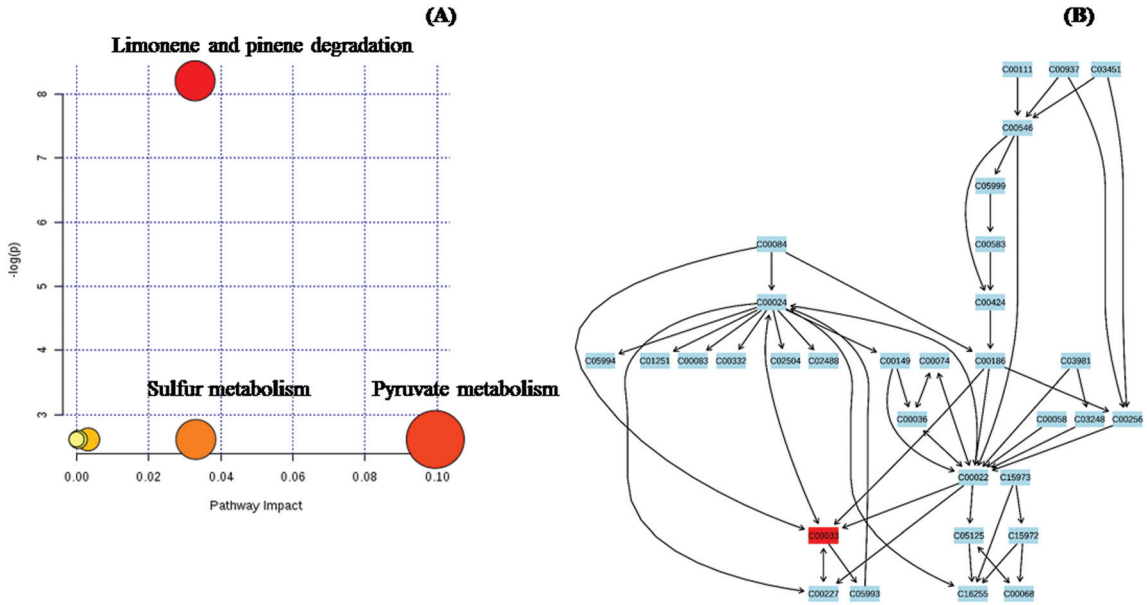

Fig. 6 (A) The metabolome view map of significant altered metabolic pathways observed in tissue samples from BC and CF groups and (B) the pyruvate metabolism. The map was generated using the reference map by KEGG (*Homo sapiens*); C00033 represents acetic acid.

is normally found in most tissues (liver, kidney, among others) and in several biofluids, namely saliva[14] and urine.[10] Moreover, limonene was also included as a significant metabolite belonging to the class of monoterpenoids involved in monooxygenase activity through cytochrome P450 and the mevalonate pathway.[32,33]

Based on the results obtained, a successful discrimination of tissue samples was achieved, according to the group showing that the volatomic tissue profile can be a useful approach to identify potential BC biomarkers.

Conclusions

This study has enabled the untargeted assessment of the metabolomic tissue profile from BC patients when compared with CF tissues using GC-qMS combined with multivariate statistical tools (PLS-DA and OPLS-DA). Twenty-nine metabolites were identified and multivariate statistical analysis revealed some metabolites significantly altered in BC patients. Of the metabolites identified, limonene, decanoic acid, acetic acid and furfural showed the highest sensitivity and specificity to discriminate BC patients. Also, the analysis of the plots leads to a metabolomic pattern comprising an array of some biochemical pathways altered in BC patients. The metabolic pathway analysis indicated that the discriminatory metabolites could originate from several dysregulated pathways in BC such as those involved in pyruvate and sulphur metabolism, and limonene degradation.

In addition, the analysis of cancer and non-cancer tissues from the same subject can also aid in balancing the effect of external interferents, such as diet or environmental exposure

and help in identifying potential biomarkers. These results suggest a possibility of identifying endogenous metabolites as a platform to discover potential BC biomarkers and pave the way to investigate the related metabolomic pathways to improve the diagnostic tools of BC.

Author contributions

Catarina Silva: Conception of studies and experimental design, BC and CF tissue extractions, sample analysis by HS-SPME/GC-qMS, data analysis and manuscript preparation. Rosa Perestrelo: Design of experiments and data analysis. Pedro Silva: Design of experiments, statistical analysis and interpretation. Filipa Capelinha: Histological classification of BC tissues. Helena Tomás: Co-supervisor of the work, conception of studies, experimental design, and manuscript preparation. José S. Câmara: Main supervisor of the work, conception of studies, experimental design, and manuscript preparation.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the FCT-Fundação para a Ciência e a Tecnologia (project PEstOE/QUI/UI0674/2019, CQM, Portuguese Government funds, and INNOINDIGO/0001/2015), the Madeira 14-20 Program (project PROEQUIPRAM – Reforço do Investimento em Equipamentos e Infraestruturas Científicas na RAM – M1420-01-0145-FEDER-000008) and by the ARDITI-Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação through the project M1420-01-0145-FEDER-000005 – Centro de Química da Madeira – CQM+ (Madeira 14-20). The authors also acknowledge the FCT for the Ph.D. grant SFRH/BD/97039/2013 and PostDoctoral fellowship SFRH/BPD/97387/2013 given to Catarina L. Silva and Rosa Perestrelo. Pedro Silva acknowledges ARDITI through the Ph.D grant under the M1420 Project – 09-5369-FSE-000001.

References

1 F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *Ca–Cancer J. Clin.*, 2018, **68**, 394–424.

2 M. J. Duffy, Serum Tumor Markers in Breast Cancer: Are They of Clinical Value?, *Clin. Chem.*, 2006, **52**, 345–351.

3 C. M. Slupsky, H. Steed, T. H. Wells, K. Dabbs, A. Schepansky, V. Capstick, W. Faught and M. B. Sawyer, Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers, *Clin. Cancer Res.*, 2010, **16**, 5835–5841.

4 L. Lavra, A. Catini, A. Ulivieri, R. Capuano, L. Baghernajad Salehi, S. Sciacchitano, A. Bartolazzi, S. Nardis, R. Paolesse, E. Martinelli, *et al.*, Investigation of VOCs associated with different characteristics of breast cancer cells, *Sci. Rep.*, 2015, **5**, 13246.

5 A. McCartney, A. Vignoli, L. Biganzoli, R. Love, L. Tenori, C. Luchinat and A. Di Leo, Metabolomics, in breast cancer: A decade in review, *Cancer Treat. Rev.*, 2018, **67**, 88–96.

6 M. M. Koek, R. H. Jellema, J. van der Greef, A. C. Tas and T. Hankemeier, Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives, *Metabolomics*, 2011, **7**, 307–328.

7 Y. Huang, Y. Li, Z. Luo and Y. Duan, Investigation of biomarkers for discriminating breast cancer cell lines from normal mammary cell lines based on VOCs analysis and metabolomics, *RSC Adv.*, 2016, **6**, 41816–41824.

8 M. M. Dougan, Y. Li, L. W. Chu, R. W. Haile, A. S. Whittemore, S. S. Han, S. C. Moore, J. N. Sampson, I. L. Andrulis, E. M. John, *et al.*, Metabolomic profiles in breast cancer:a pilot case-control study in the breast cancer family registry, *BMC Cancer*, 2018, **18**, 532.

9 C. Denkert, E. Bucher, M. Hilvo, R. Salek, M. Orešič, J. Griffin, S. Brockmöller, F. Klauschen, S. Loibl, D. K. Barupal, *et al.*, Metabolomics of human breast cancer: new approaches for tumor typing and biomarker discovery, *Genome Med.*, 2012, **4**, 37.

10 C. L. Silva, M. Passos and J. S. Câmara, Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers–a powerful strategy for breast cancer diagnosis, *Talanta*, 2012, **89**, 360–368.

11 C. L. Silva, R. Perestrelo, P. Silva, H. Tomás and J. S. Câmara, Volatile metabolomic signature of human breast cancer cell lines, *Sci. Rep.*, 2017, **7**, 43969.

12 C. L. Silva, M. Passos and J. S. Câmara, Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry, *Br. J. Cancer*, 2011, **105**, 1894–1904.

13 P. Porto-Figueira, J. A. M. Pereira and J. S. Câmara, Exploring the potential of needle trap microextraction combined with chromatographic and statistical data to discriminate different types of cancer based on urinary volatomic biosignature, *Anal. Chim. Acta*, 2018, **1023**, 53–63.

14 C. Cavaco, J. A. M. Pereira, K. Taunk, R. Taware, S. Rapole, H. Nagarajaram and J. S. Câmara, Screening of salivary volatiles for putative breast cancer discrimination: an exploratory study involving geographically distant populations, *Anal. Bioanal. Chem.*, 2018, **410**, 1–10.

15 J. Budczies, B. M. Pfitzner, B. Györffy, K.-J. Winzer, C. Radke, M. Dietel, O. Fiehn and C. Denkert, Glutamate enrichment as new diagnostic opportunity in breast cancer, *Int. J. Cancer*, 2015, **136**, 1619–1628.

16 J. Budczies, S. F. Brockmöller, B. M. Müller, D. K. Barupal, C. Richter-Ehrenstein, A. Kleine-Tebbe, J. L. Griffin, M. Orešič, M. Dietel, C. Denkert, *et al.*, Comparative metabolomics of estrogen receptor positive and estrogen receptor negative breast cancer: alterations in glutamine and beta-alanine metabolism, *J. Proteomics*, 2013, **94**, 279–288.

17 J. Xia, D. S. Wishart, J. Xia and D. S. Wishart, Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis, in *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016, pp. 14.10.1–14.10.91. ISBN: 9780471250951.

18 W. Filipiak, A. Filipiak, A. Sponring, T. Schmid, B. Zelger, C. Ager, E. Klodzinska, H. Denz, A. Pizzini, P. Lucciarini, *et al.*, Comparative analyses of volatile organic compounds (VOCs) from patients, tumors and transformed cell lines for the validation of lung cancer-derived breath markers, *J. Breath Res.*, 2014, **8**, 027111.

19 D. Poli, P. Carbognani, M. Corradi, M. Goldoni, O. Acampa, B. Balbi, L. Bianchi, M. Rusca and A. Mutti, Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study, *Respir. Res.*, 2005, **6**, 71.

20 I. Ahmed, R. Greenwood, B. Costello, N. Ratcliffe and C. S. Probert, Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease, *Aliment. Pharmacol. Ther.*, 2016, **43**, 596–611.

21 P. Mochalski, M. Leja, E. Gasenko, R. Skapars, D. Santare, A. Sivins, D. E. Aronsson, C. Ager, C. Jaeschke, G. Shani, *et al.*, Ex vivo emission of volatile organic compounds from gastric cancer and non-cancerous tissue, *J. Breath Res.*, 2018, **12**, 046005.

22 M. Raman, I. Ahmed, P. M. Gillevet, C. S. Probert, N. M. Ratcliffe, S. Smith, R. Greenwood, M. Sikaroodi, V. Lam, P. Crotty, *et al.*, Fecal Microbiome and Volatile Organic Compound Metabolome in Obese Humans With Nonalcoholic Fatty Liver Disease, *Clin. Gastroenterol. Hepatol.*, 2013, **11**, 868–875.e3.

23 J. Carrola, C. M. Rocha, A. S. Barros, A. M. Gil, B. J. Goodfellow, I. M. Carreira, J. Bernardo, A. Gomes, V. Sousa, L. Carvalho, *et al.*, Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine, *J. Proteome Res.*, 2011, **10**, 221–230.

24 L. Wang, Y. Tang, S. Liu, S. Mao, Y. Ling, D. Liu, X. He and X. Wang, Metabonomic profiling of serum and urine by (1) H NMR-based spectroscopy discriminates patients with chronic obstructive pulmonary disease and healthy individuals, *PLoS One*, 2013, **8**, e65675.

25 V. M. Asiago, L. Z. Alvarado, N. Shanaiah, G. A. N. Gowda, K. Owusu-Sarfo, R. A. Ballas and D. Raftery, Early Detection of Recurrent Breast Cancer Using Metabolite Profiling, *Cancer Res.*, 2010, **70**, 8309–8318.

26 S. Srivastava, R. Roy, S. Singh, P. Kumar, D. Dalela, S. N. Sankhwar, A. Goel and A. A. Sonkar, Taurine – a possible fingerprint biomarker in non-muscle invasive bladder cancer: A pilot study by 1H NMR spectroscopy, *Cancer Biomarkers*, 2010, **6**, 11–20.

27 J. Jung, Y. Jung, E. J. Bang, S. Cho, Y.-J. Jang, J.-M. Kwak, D. H. Ryu, S. Park and G.-S. Hwang, Noninvasive Diagnosis and Evaluation of Curative Surgery for Gastric Cancer by Using NMR-based Metabolomic Profiling, *Ann. Surg. Oncol.*, 2014, **21**, 736–742.

28 M. Bylesjö, M. Rantalainen, O. Cloarec, J. K. Nicholson, E. Holmes and J. Trygg, OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, *J. Chemom.*, 2006, **20**, 341–351.

29 S. G. Baker, The Central Role of Receiver Operating Characteristic (ROC) Curves in Evaluating Tests for the Early Detection of Cancer, *JNCI, J. Natl. Cancer Inst.*, 2003, **95**, 511–515.

30 K. Hajian-Tilaki, Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation, *Caspian J. Intern. Med.*, 2013, **4**, 627–635.

31 J. Xia, D. I. Broadhurst, M. Wilson and D. S. Wishart, Translational biomarker discovery in clinical metabolomics: an introductory tutorial, *Metabolomics*, 2013, **9**, 280–299.

32 S. Janocha, D. Schmitz and R. Bernhardt, *Terpene Hydroxylation with Microbial Cytochrome P450 Monooxygenases*, 2015, pp. 215–250.

33 Z. Costello and H. G. Martin, A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data, *npj Syst. Biol. Appl.*, 2018, **4**, 19.