

# Feasibility of Detecting Prostate Cancer by Ultrapformance Liquid Chromatography–Mass Spectrometry Serum Metabolomics

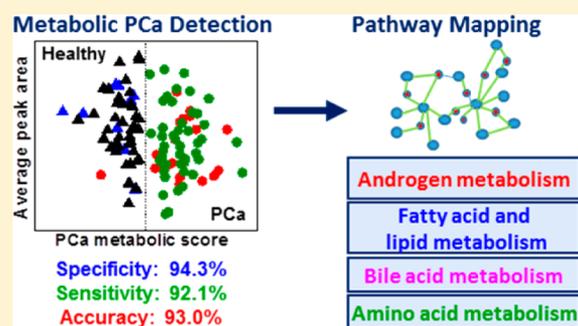
Xiaoling Zang,<sup>†,⊥</sup> Christina M. Jones,<sup>†,⊥</sup> Tran Q. Long,<sup>‡</sup> María Eugenia Monge,<sup>†,#</sup> Manshui Zhou,<sup>†</sup> L. DeEte Walker,<sup>§</sup> Roman Mezencev,<sup>§</sup> Alexander Gray,<sup>‡</sup> John F. McDonald,<sup>§,||</sup> and Facundo M. Fernández<sup>\*,†,||</sup>

<sup>†</sup>School of Chemistry and Biochemistry, <sup>‡</sup>College of Computing, <sup>§</sup>School of Biology, Integrated Cancer Research Center, and <sup>||</sup>Parker H. Petit Institute of Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States

## S Supporting Information

**ABSTRACT:** Prostate cancer (PCa) is the second leading cause of cancer-related mortality in men. The prevalent diagnosis method is based on the serum prostate-specific antigen (PSA) screening test, which suffers from low specificity, overdiagnosis, and overtreatment. In this work, untargeted metabolomic profiling of age-matched serum samples from prostate cancer patients and healthy individuals was performed using ultrapformance liquid chromatography coupled to high-resolution tandem mass spectrometry (UPLC-MS/MS) and machine learning methods. A metabolite-based *in vitro* diagnostic multivariate index assay (IVDMIA) was developed to predict the presence of PCa in serum samples with high classification sensitivity, specificity, and accuracy. A panel of 40 metabolic spectral features was found to be differential with 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. The performance of the IVDMIA was higher than the prevalent PSA test. Within the discriminant panel, 31 metabolites were identified by MS and MS/MS, with 10 further confirmed chromatographically by standards. Numerous discriminant metabolites were mapped in the steroid hormone biosynthesis pathway. The identification of fatty acids, amino acids, lysophospholipids, and bile acids provided further insights into the metabolic alterations associated with the disease. With additional work, the results presented here show great potential toward implementation in clinical settings.

**KEYWORDS:** prostate cancer, prostate cancer detection, untargeted metabolomics, oncometabolomics, ultrapformance liquid chromatography, mass spectrometry, machine learning methods, support vector machines, *in vitro* diagnostic multivariate index assay, IVDMIA



## INTRODUCTION

Prostate cancer (PCa) is the second leading cause of cancer-related mortality in men worldwide, with 30 000 deaths per year in the U.S. alone.<sup>1</sup> The prevalent diagnosis method is based on the triad of digital rectal examination, blood prostate-specific antigen (PSA) measurement, and ultrasound-guided prostate biopsy. Although the introduction of PSA screening decreased mortality by 4% between 1994 and 2006,<sup>2</sup> the use of PSA as a diagnostic serum marker still presents several drawbacks. The concentration of this protein in the blood-stream increases during the development of cancer but also can be secreted as a result of benign prostatic hyperplasia, prostatitis, or other traumas to prostate cells.<sup>3</sup> Therefore, this method suffers from low specificity and consequent overdiagnosis and overtreatment.<sup>4–7</sup> Moreover, approximately 15% of patients with PCa have PSA values lower than the commonly used cutoff point of 4.0 ng mL<sup>-1</sup>, leaving many cases undetected.<sup>8</sup> These issues have led to an increased interest in using untargeted metabolite profiling to discover new differential metabolic biomarkers that could improve the specificity

of PCa diagnosis.<sup>9</sup> Metabolic biomarkers are used as a routine tool in screening newborns for the presence of inborn errors of metabolism by means of tandem mass spectrometry;<sup>10,11</sup> however, global metabolite profiling of PCa patients still remains at an early stage, and there is no biomarker panel currently in use for clinical testing.<sup>9</sup>

Current research has shown some evidence of metabolic alterations associated with PCa. Tissue sarcosine levels have been suggested as a potential biomarker for the aggressive form of the disease in a metabolomic profiling study using both liquid and gas chromatography coupled to mass spectrometry (LC-MS and GC-MS).<sup>12</sup> Its concentration in prostate-cancer-related tissue specimens was highly increased during PCa progression to metastasis, but differences in urine were much less marked.<sup>12</sup> These results have been very prominent but somewhat controversial as other targeted studies failed in the attempt to differentiate healthy individuals from cancer patients

Received: April 21, 2014

Published: June 12, 2014

based on sarcosine concentration in biological fluids and cancerous tissues.<sup>9,13–15</sup> The analysis of cancerous tissues by proton high-resolution magic-angle spinning nuclear magnetic resonance (NMR) spectroscopy has shown a decrease in the concentrations of citrate and polyamines and increases in cholines, glycerophospholipids, and lactate concentrations during PCa proliferation.<sup>16,17</sup> Increased levels of cholesterol as well as alterations in amino acid metabolism were detected in metastatic bone samples by GC-MS.<sup>18</sup> However, all of these studies included too few patients to offer strong leads on the metabolic alterations associated with PCa. A panel of plasma lipids that included phosphatidylethanolamines, ether-linked phosphatidylethanolamines, and ether-linked phosphatidylcholines was proposed to discriminate PCa patients from healthy groups through direct infusion electrospray ionization tandem MS.<sup>19</sup> The authors demonstrated that a combination of multiple biomarkers with multivariate analysis and various classification algorithms yielded better predictive power for the diagnosis of PCa than univariate analysis of single lipid species. However, the predictive power was not compared with that of PSA, as this information was not available at the time of cohort design.<sup>19</sup> More robust metabolic models still need to be developed for improved understanding of disease progression and more reliable PCa detection.

From a statistical point of view, analysis of metabolomic data sets represents a significant challenge, and robust approaches are necessary to handle, extract, and classify the relevant information from the vast amount of data generated.<sup>20</sup> Along these lines, the application of machine learning algorithms capable of identifying differentiating metabolites has seen increased interest.<sup>21,22</sup> For example, a mass spectrometric platform in combination with support vector machines (SVMs) has shown some promise for the detection of ovarian cancer in blood sera,<sup>23,24</sup> for lung cancer in blood plasma,<sup>25</sup> and for breast cancer in urine,<sup>26</sup> to mention a few examples.

In this study, ultraperformance liquid chromatography coupled to high-resolution mass spectrometry and tandem mass spectrometry (UPLC-MS and MS/MS) combined with machine learning methods was used to profile and identify a panel of metabolites in blood sera that discriminates PCa patients from healthy individuals. Based on these, we have developed an *in vitro* diagnostic multivariate index assay (IVDMIA)<sup>27</sup> that provides a score predicting the presence or absence of PCa. Moreover, we have chemically identified 10 discriminant metabolites and putatively identified another 21 metabolic spectral features from the discriminant panel.

## MATERIALS AND METHODS

### Cohort Description

Age-matched blood serum samples were obtained from 64 PCa patients (age range 49–65, mean age  $59 \pm 4$  years) and 50 healthy individuals (age range 45–76, mean age  $57 \pm 7$  years). At the 0.05 level, the population means were not significantly different with the two-sample *t* test. The cohort ethnicity was as follows: 28 African American (24.6%); 76 Caucasian (66.7%); 5 Hispanic (4.4%); 2 Asian (1.8%); 2 Jewish ancestry (1.8%); and 1 unknown (0.9%). After approval by the Institutional Review Board (IRB), blood samples were collected at Saint Joseph's Hospital of Atlanta (GA, USA) by venipuncture from each donor into evacuated blood collection tubes that contained no anticoagulant. Serum was obtained by centrifugation at 5000 rpm for 5 min at 4 °C. Immediately after centrifugation, 200  $\mu$ L

aliquots of serum were frozen and stored at  $-80$  °C for further use. The sample collection and storage procedures for PCa patients and healthy individuals were identical. Gleason scores were obtained for 61 PCa patients.

### Chemicals

Healthy human blood serum (S7023-50 mL) and acetic acid ( $\geq 99.7\%$ ) were purchased from Sigma-Aldrich Corp. (St. Louis, MO, USA). Omnisolv LC-MS grade acetonitrile, Omnisolv high-purity dichloromethane, and HPLC grade acetone were purchased from EMD (Billerica, MA, USA). LC-MS grade methanol and 2-propanol were purchased from J.T. Baker Avantor Performance Materials, Inc. (Center Valley, PA, USA). Ultrapure water with 18.2 M $\Omega$ ·cm resistivity (Barnstead Nanopure UV ultrapure water system, USA) was used to prepare mobile phases. Uric acid ( $\geq 99\%$ ), azelaic acid (98%), undecanedioic acid (97%), heptadecanoic acid ( $\geq 98\%$ ), and decanoic acid ( $\geq 98\%$ ) were purchased from Sigma-Aldrich Corp. (St. Louis, MO, USA). Hexadecanedioic acid (98%) was purchased from Ark Pharm, Inc. (Libertyville, IL, USA). Phenylalanyl phenylalanine was purchased from MP Biomedicals (Solon, OH, USA). Phenylacetyl glutamine was purchased from Bachem (Hauptstrasse, Bubendorf, Switzerland). Indoxyl sulfate potassium was purchased from Alfa Aesar (Ward Hill, MA, USA). 1-Stearoyl-2-hydroxy-*sn*-glycero-3-phosphocholine/lysoPC (18:0/0:0) was purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA).

### Sample Preparation and Ultraperformance Liquid Chromatography–Mass Spectrometry (UPLC-MS)

A stock sample of healthy human blood serum was used to develop the serum sample preparation protocol and UPLC-MS method. Serum samples were thawed on ice, and protein precipitation was performed by the addition of a mixture of acetone, acetonitrile, and methanol (1:1:1 v/v) to 100  $\mu$ L of serum in a 3:1 volume ratio. Samples were vortex-mixed for 20 s and centrifuged at 16 000g for 5 min. After centrifugation, 800  $\mu$ L of dichloromethane was added to 350  $\mu$ L of supernatant and vortex-mixed. Following the addition of 250  $\mu$ L of deionized water, samples were vortex-mixed again to extract the nonpolar lipid fraction. The aqueous phase was used for metabolite analysis by UPLC-MS. Samples were randomly separated into seven batches and consecutively analyzed. The instrument was calibrated before analysis and solvent, and sample preparation blanks were jointly analyzed with the samples in a random order.

UPLC-MS analysis was performed using a Waters ACQUITY Ultra Performance LC (Waters Corporation, Manchester, UK) system, fitted with a Waters ACQUITY UPLC BEH C<sub>18</sub> column (2.1  $\times$  50 mm, 1.7  $\mu$ m particle size), and coupled to a high-resolution accurate mass (HRAM) Synapt G2 high-definition mass spectrometry (HDMS) system (Waters Corporation, Manchester, UK). The Synapt G2 HDMS is a hybrid quadrupole-ion mobility-orthogonal acceleration time-of-flight instrument with typical resolving power of 20 000 fwhm and mass accuracy of 9 ppm at *m/z* 554.2615. The instrument was operated in negative ion mode with a probe capillary voltage of 2.3 kV and a sampling cone voltage of 45 V. The source and desolvation temperatures were 120 and 350 °C, respectively, and the nitrogen desolvation flow rate was 650 L h<sup>-1</sup>. The mass spectrometer was calibrated across the range of *m/z* 50–1800 using a 0.5 mM sodium formate solution prepared in 90:10 2-propanol/water v/v. Data were mass corrected during acquisition using a leucine

enkephalin reference spray (LockSpray) infused at  $2 \mu\text{L min}^{-1}$ . Data were acquired in the 50–1750  $m/z$  range, and the scan time was set to 1 s. Data acquisition and processing were carried out using MassLynx v4.1. The chromatographic method for sample analysis involved elution with acetonitrile (mobile phase A) and water with 0.1% acetic acid (mobile phase B) using the following gradient program: 0–1 min 0–10% A; 1–2.5 min 10–15% A; 2.5–4 min 15–22% A; 4–6 min 22–38% A; 6–9 min 38–65% A; 9–12 min 65–80% A; 12–16 min 80–100% A; 16–18 min 100% A. The flow rate was constant at  $0.25 \text{ mL min}^{-1}$  for 12 min. It was increased to  $0.30 \text{ mL min}^{-1}$  between 12 and 16 min and from  $0.30$  to  $0.45 \text{ mL min}^{-1}$  between 16 and 18 min. The gradient was returned to its initial conditions over a period of 8 min after each sample injection. The column temperature was set to  $35^\circ\text{C}$ , the autosampler tray temperature was set to  $5^\circ\text{C}$ , and the injection volume was  $10 \mu\text{L}$ . UPLC-MS/MS experiments were performed by acquiring mass spectra with applied voltages between 5 and 50 V in the trap cell, using ultrapurity argon ( $\geq 99.999\%$ ) as the collision gas.

### Data Analysis

After UPLC-MS analysis, metabolic features (retention time ( $R_t$ ),  $m/z$  pairs) were extracted from chromatograms using MarkerLynx XS software. This procedure involved chromatogram alignment, peak picking and integration, peak area extraction, and normalization. The matrix containing sample peak areas for each feature ( $R_t$ ,  $m/z$ ) was utilized to build a model for sample classification and to find the minimum set of discriminant features by means of linear support vector machines (SVMs).<sup>28</sup> This supervised classification technique is effective at handling high dimensionality data as those produced in the present work. For a binary classification problem, linearly separable samples represented as a row vector  $\mathbf{x}$  had membership of two classes  $g$  ( $= \text{H or D}$ ), where H stands for healthy and D for PCa disease with labels  $c = -1$  for class H and  $+1$  for class D. In order to build the classification model, 70% of the samples were randomly selected as a training set and 30% as a test set. Within the training set, 10% of samples were used for validation and to find the minimum set of discriminant features that maximized accuracy in the classification through a recursive feature elimination (RFE) method.<sup>23</sup> The decision function that separated the two classes, defined here as the IVDMA “PCa metabolic score”, was as follows:

$$\text{PCa metabolic score} = b + \sum_{j=1}^j w_j x_{ij} \quad (1)$$

$$g(\mathbf{x}_i) = \text{sgn}(w\mathbf{x}'_i + b) = \text{sgn}(\text{PCa metabolic score}) \quad (2)$$

where  $w$  and  $b$  are the weight and bias parameters that were determined from the training set and  $j$  is the total number of features. The sign of the PCa metabolic score determined to which class a sample was assigned: class H if negative and class D if positive. In this classification function, the two classes were divided in the dataspace by a hyperplane  $w\mathbf{x}' + b = 0$  that maximized the margins between samples of different classes. The margin between the two classes was defined such that

$$w\mathbf{x}' + b \geq 1, \quad c = +1 \quad (3)$$

$$w\mathbf{x}' + b \leq -1, \quad c = -1 \quad (4)$$

To estimate the classification and feature selection performance, 10 iterative validations were performed to randomly select

the training and test sets. The statistical significance of the model was further assessed through hypothesis testing by permutation tests. No assumptions were made in this nonparametric approach to hypothesis testing regarding the data distribution, and the  $p$  value was computed as the cumulative sum using the empirical distribution. Two permutation tests were performed using 100 permutation samples with the following null hypothesis:

- (i) Null hypothesis 1: feature and labels (positive/negative) are independent (i.e., indifference when class labels are permuted).
- (ii) Null hypothesis 2: features are independent within each class (i.e., indifference when value of each features are permuted within each class).

If the  $p$  value  $< \alpha$  ( $\alpha = 0.05$ ), the null hypothesis  $H_0$  was rejected; otherwise, the observed result was not statistically significant.

Additionally, principal component analysis (PCA) was used to evaluate the performance of all extracted metabolic features or subsets of them in an unsupervised manner with MATLAB R2011b (version 7.13.0, The MathWorks, Inc., Natick, MA, USA) and the PLS Toolbox (v.6.71, Eigenvector Research, Inc., Wenatchee, WA, USA). Data were preprocessed by autoscaling.

### Metabolite Identification Procedure

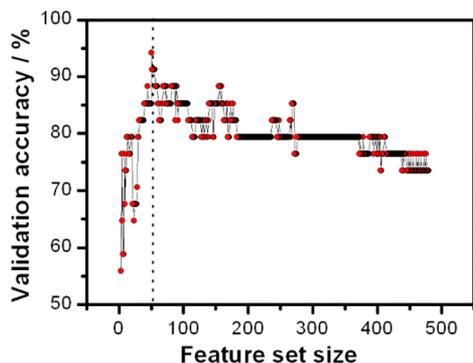
Compound identification was attempted for the 40 discriminant features remaining after the feature selection processes. Due to the biological complexity of serum samples, adduct ion analysis was first performed to ensure the unambiguous assignment of the signal of interest in each mass spectrum. Adduct ions corresponding to SVM-selected variables that were investigated in the mass spectra included  $[\text{M} - \text{H}]^-$ ,  $[\text{M} + \text{Cl}]^-$ ,  $[\text{M} + \text{Br}]^-$ ,  $[\text{M} + \text{CH}_3\text{COO}]^-$ ,  $[\text{M} + \text{HCOO}]^-$ ,  $[\text{M} + \text{CF}_3\text{COO}]^-$ ,  $[\text{M} + \text{Na} - 2\text{H}]^-$ ,  $[\text{M} + \text{K} - 2\text{H}]^-$ ,  $[\text{M} - \text{H}_2\text{O} - \text{H}]^-$ ,  $[\text{M} + \text{H}_2\text{O} - \text{H}]^-$  species, which are typically observed in negative ion mode electrospray ionization. The expected  $m/z$  values for common adduct species were calculated and compared with the experimental values from peaks within the spectra. For spectra in which no confirmatory adducts were present, the accurate mass of the candidate neutral molecule was calculated based on the assumption that the peak of interest corresponded to  $[\text{M} - \text{H}]^-$ . Elemental formulas were generated based on the mass accuracy of the peak of interest and isotopic patterns with a mass error of 8 mDa, using MassLynx 4.1. The list of elements included in the search was C, H, N, O, P, S, Cl, and Br. The list of generated elemental formulas was searched against the Metlin database,<sup>29</sup> HMDB,<sup>30</sup> and in order to determine the possible endogenous metabolite candidates. The MS/MS Metlin and MassBank<sup>31</sup> databases and a literature survey were subsequently used to confirm the identity of putative candidates. Fragmentation patterns were also manually analyzed to discriminate between different isobaric species. Additionally, commercially available chemical standards were analyzed by UPLC-MS and MS/MS to confirm tentative metabolite identities by retention time and mass spectral matching.

## RESULTS AND DISCUSSION

### Classification Performance

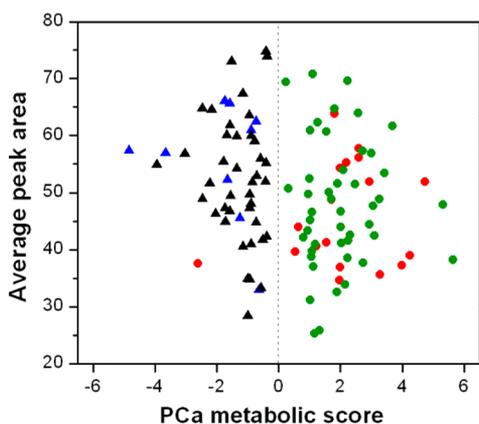
UPLC-MS analysis in negative ion mode allowed the interrogation of highly complex serum samples from PCa patients and healthy individuals, revealing a total of 480 features

( $R_v$ ,  $m/z$  pairs). The extracted features were used to build a discriminant SVM model for sample classification. An optimum set of 51 discriminant features was found to maximize classification accuracy through a recursive feature elimination method,<sup>23</sup> as illustrated in Figure 1. Out of the 51 selected



**Figure 1.** Evolution of classification accuracy for a validation sample subset consisting of 10% of the training samples as a function of the number of features retained. The minimum discriminant feature set that maximizes classification accuracy is highlighted with a dashed line.

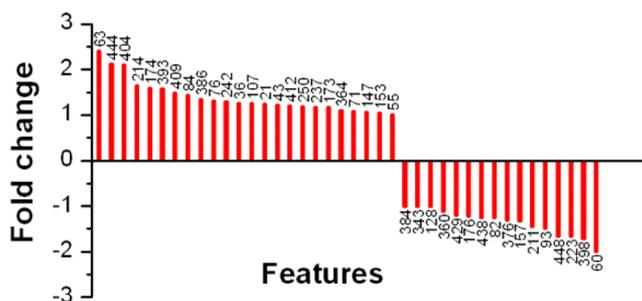
features, seven were found to be only present in less than 2% of the samples; two features were identified as acetaminophen and its sulfite adduct, and two additional features were identified as adducts or fragments of other features in the subset and were thus removed from further consideration. The optimum panel that best discriminated PCa patients from healthy individuals was thus reduced to 40 features, demonstrating that the feature selection process accomplished a high reduction in problem dimensionality. Figure 2 illustrates the “PCa metabolic scores”



**Figure 2.** Visualization of the PCa metabolic scores obtained by SVMs in 1 out of 10 iterative model validations based on 40 discriminant features. Green circles correspond to PCa patients in the training set; black triangles correspond to controls in the training set; red circles correspond to PCa patients in the test set built for the iteration shown, and blue triangles correspond to healthy individuals in the test set. The dotted line shows the projection of the separating hyperplane:  $w\mathbf{x}' + b = 0$ .

obtained for the training and the test sets of randomly selected samples that were used to construct and evaluate the classification model, respectively. The separation of the two sample classes (H or D) was determined in the data space by the optimal separating hyperplane for which the margin between the most similar samples in each group was largest,

illustrated with a dotted line in the figure. The samples with scores equal to 1 or  $-1$  are the support vectors of the model. For the particular cross-validation iteration illustrated in Figure 2, only one sample was misclassified as a false negative. The remaining nine iterative validations with their respective training and test sets are illustrated in Figure S1 (Supporting Information). Based on these 40 discriminant features, serum samples were successfully classified as cancerous or healthy with 93.0% accuracy, 92.1% sensitivity, and 94.3% specificity. These values were calculated as the averages from 10 distinct test sets, illustrated in Figures 2 and S1. In addition, the statistical significance of the model was further evaluated through hypothesis testing, and at 0.05 significance level, the null hypothesis was rejected for all permutations generated ( $p$  value = 0.0099). Unambiguously, the classifier did not yield a better leave-one-out-cross-validation (LOOCV) accuracy rate than the original data. These results suggest a promising approach that could form the basis for a PCa IVDMA. In particular, of the 40 differential features, 24 were found to increase in sera from PCa patients, and 16 were found to decrease in PCa, as illustrated in Figure 3. It is important to underscore, however,



**Figure 3.** Fold change of average peak areas of each discriminant feature. Positive fold changes are calculated as the ratio of average peak areas between PCa patients and healthy individuals, and negative fold changes are calculated as the negative ratio of average peak areas between healthy individuals and PCa patients. Features are labeled with their codes.

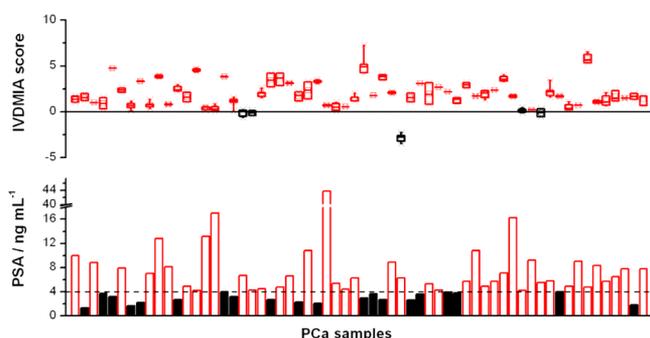
that the strength of this IVDMA resides in the combination of multiple metabolic features using an interpretation function to yield a single, patient-specific result to be used in the disease diagnosis and not on the average fold change of each differential feature.

In order to evaluate the possible risk of data overfitting by SVMs,<sup>28</sup> a simple unsupervised approach was also used to examine the data set. PCA score plots were generated for both the 40 discriminant features set obtained by SVMs and the starting set of 480 features. Figure S2 (Supporting Information) shows the results for each case. Using the best 40 features, three principal components containing 33.6% of the total variance provided a good degree of separation between classes, as illustrated in Figure S2a. Sample separation in the PCA score plot was mainly achieved by the contribution of PC3. Loadings for PC3 are displayed in Figure S2b. Interestingly, Figure S2c shows that PCA does not provide any distinguishable clustering when applied to the initial set of 480 features, further supporting the use of RFE and SVMs when handling high dimensionality data volumes as those in the present work. Given the clustering observed in PCA when using the 40 discriminant feature subset, the risk of the high classification

accuracy of SVM models being a product of overfitting is greatly diminished.

### IVDMIA versus PSA Diagnosis

The Gleason scores for the PCa patients, summarized in Table S1 (Supporting Information), indicate that the most common tumor patterns presented by the patients derived from moderate to aggressive cancers. However, the PSA test performed at surgery did not follow this histological evidence for the entire PCa cohort, as 33% of patients with PCa ( $n = 20$ ) had PSA values lower than the commonly used cutoff point of  $4.0 \text{ ng mL}^{-1}$ . Figure 4 compares PSA and IVDMIA results in



**Figure 4.** Comparison of IVDMIA vs PSA diagnosis performance for 62 PCa patients. True positive and false negative outputs are highlighted in red and black, respectively. The cutoff point of  $4.0 \text{ ng mL}^{-1}$  used in PSA-based diagnosis is indicated with a dotted line. The IVDMIA score output is presented as box plots in the figure, each of which is generated by results obtained for each of the 10 test sets where each sample was selected for validation. No comparison is shown for two of the 64 PCa samples as they were not randomly selected in any of the 10 cross-validation iterations.

terms of true positive and false negative outputs, highlighted in red and black, respectively. The IVDMIA outputs provided by the randomly selected 10 test sets are visualized as box plots in the figure and show that the IVDMIA was able to correctly predict 100% of the true positives that were incorrectly diagnosed as negatives by the PSA test. The false negative results provided by the IVDMIA derived from one sample that was misclassified in all test sets and four samples that were misclassified in at least one test set. The classification performance obtained with this cohort shows promise toward prostate cancers that would go undetected by the PSA method.

The use of multiple discriminant features by this metabolic IVDMIA yields higher predictive power for PCa diagnosis than the univariate analysis of a single marker such as with the PSA method.

### IVDMIA Potential in Clinical Applications

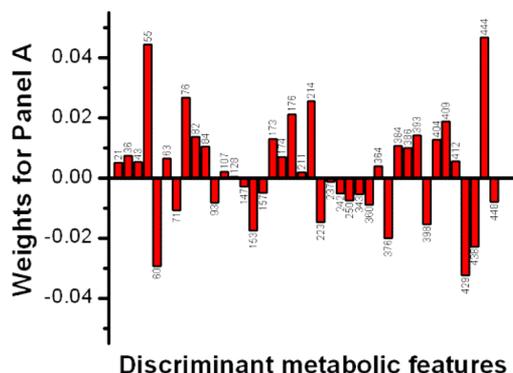
To determine the fraction of samples in which the discriminant features were present, and to evaluate the feasibility of implementing the PCa IVDMIA in clinical laboratory settings through targeted triple quadrupole mass spectrometry-based assays, smaller subgroups of the optimum 40 discriminant features, subsequently referred to as “panel A”, were investigated (Table 1). These subpanels were chosen to provide the minimum number of features that collectively captured metabolic PCa patterns with a high level of accuracy, specificity, and sensitivity. The selection of these additional subpanels was based on the fraction of features that were present in 50, 70, or 90% of the entire sample cohort, in either PCa patients or healthy controls. Table 1 summarizes the different panels constructed following these criteria, with their corresponding subset of discriminant features. These panels were used to build new SVM models and cross-validated to provide average values of accuracy, specificity, and sensitivity from 10 independent randomly selected training and testing sets. Thirty-eight out of 40 discriminant features were present in more than 50% of healthy controls (panel B), and 35 out of 40 were present in more than 50% of PCa samples (panel C), providing similar accuracy, specificity, and sensitivity as panel A. When the criterion for feature presence was made more stringent, from panel A to panel G, the accuracy, specificity, and sensitivity decreased by only  $\sim 10\%$ , suggesting the robust biological role that the detected features might have. In other words, the different feature subpanels were not highly sensitive to a reduction in the number of discriminant features, suggesting that the smaller number of metabolites contained in subpanel G could still be potentially used to build a more focused, simpler IVDMIA for PCa detection in a clinical setting. To further test this finding, another SVM model was created with only those 13 features that could be confidently assigned to metabolites in subpanel G by HRAM MS and MS/MS (Table S2, Supporting Information). It was found that this model still provided high classification sensitivity (88.3%), specificity (80.3%), and accuracy (85.0%). The mass spectrometric assay for such model would be much simpler to implement in a targeted fashion due to the reduced number of transitions that a UPLC-MS/MS triple quadrupole method

**Table 1.** Discriminant Feature (Sub)panels for PCa Detection

panel	accuracy	specificity	sensitivity	discriminant features (#)	discriminant feature codes	% healthy samples, % PCa samples
A	93.0	94.3	92.1	40	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153, 21, 364, 404, 242, 237	>0%, >0%
B	91.2	90.6	91.7	38	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153, 21, 364, 404	>50%, >0%
C	90.2	87.2	91.8	35	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153	>50%, >50% and >0%, >50%
D	86.1	87.2	85.3	28	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214	>0%, >70%
E	84.4	80.0	85.8	25	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384	>70%, >70% and >70%, >0%
F	85.0	80.0	88.8	22	147, 36, 71, 60, 55, 409, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384	>90%, >0%
G	80.0	81.0	79.3	17	147, 36, 71, 60, 55, 409, 386, 438, 157, 176, 82, 393, 173, 84, 343, 429, 384	>90%, >90% and >0%, >90%

would require, allowing higher analysis throughput and minimizing cost.

The set of 40 SVM weights obtained for panel A from the optimal classification model are shown in Figure 5. The figure



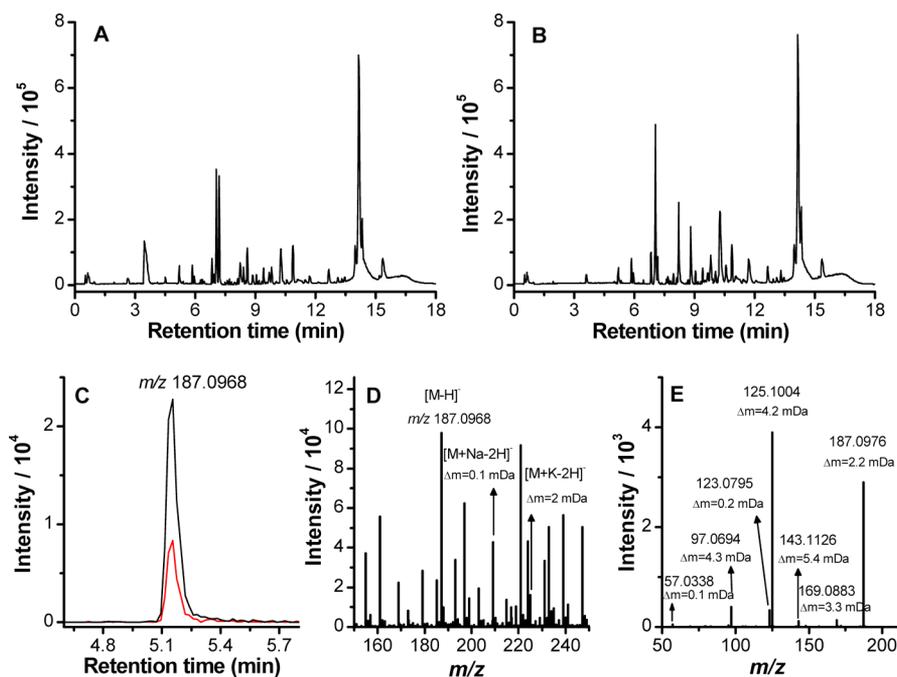
**Figure 5.** Weights for the 40 discriminant metabolic features in panel A. Metabolic features are labeled with their codes.

shows the individual contribution of each of the 40 discriminant metabolic features in the computed PCa metabolic score, that is, the weight of each discriminant metabolite in the classification. It is interesting to note that some features with high weights in the SVM model, such as feature 60, 444, 409, or 429, also have large absolute values in the PC3 loadings plot (Figure S2b, Supporting Information). Figure S3 (Supporting Information) shows a comparison of the different sets of weights for the different panels described in Table 1, sorted from the largest to lowest value in panel A and expanded to panels B–G. The figure shows that the sign of the weights

generally remained the same across the panels, in agreement with the fact that accuracy, specificity, and sensitivity were highly conserved even after restricting the presence of discriminant features to those present in a majority of the patients within the cohort. It was seen that for the most restrictive panels, those features with weights equal to zero (i.e., those that do not contribute to the panels) are those with lower weights in panel A.

#### Identification of Metabolites Used in the IVDMA

Once the robustness of the model was established, chemical identification of the 40 discriminant metabolic features was attempted. Figure 6 exemplifies the procedure utilized for identification of feature 60. Figure 6A,B shows the different base peak intensity chromatograms (BPI) obtained for serum samples of a typical PCa patient and healthy individual. As differences between metabolomes and the corresponding features in the BPI chromatograms arise both from the presence of the disease and from differences in diet, lifestyle, and numerous other factors,<sup>32</sup> chemical identification of endogenous metabolites was attempted only for the 40 discriminant metabolic features. The high resolving power of the time-of-flight analyzer used allowed generating highly selective extracted ion chromatograms for each discriminant feature, as illustrated in Figure 6C. Adduct ion analysis (Figure 6D) was used to ensure the unambiguous assignment of the signal of interest in the electrospray ionization mass spectrum, and the isotopic pattern and accurate masses were used to generate a list of possible candidate elemental formulas that were searched against databases. Moreover, UPLC-MS/MS experiments were performed to confirm the identities of these candidate metabolites responsible for classification. Tandem MS spectra were compared to those in databases or literature,



**Figure 6.** Base peak intensity chromatograms obtained for typical serum samples from a patient with PCa (A) and a healthy individual (B). (C) Extracted ion chromatogram for  $m/z$  187.0968  $\pm$  0.0050 generated from a PCa patient sample (red line) and a healthy individual (black line). These were generated from the data shown in A and B, respectively. (D) Adduct ion analysis for discriminant feature at  $m/z$  187.0968. Mass errors are calculated with respect to the theoretical values for azelaic acid ( $C_9H_{16}O_4$ ). (E) Tandem MS spectrum for the  $m/z$  187.0968 precursor ion using a collision cell voltage of 15 V. The matching of tandem MS fragmentation patterns between the experimental spectrum and the metabolite candidate is illustrated by the mass errors calculated as differences with the values in the Metlin database.

Table 2. Results for the Chemical Identification Workflow for Various Discriminant Features<sup>a</sup>

feature code	retention time (min)	m/z	ion type	elemental formula	theoretical m/z	$\Delta m$ (mDa)	tentative metabolite identification	ref	panel
60	5.10	187.0970	[M - H] <sup>-</sup>	C <sub>9</sub> H <sub>16</sub> O <sub>4</sub>	187.0970	0.0	<b>nonanedioic acid (azelaic acid)</b>	50	G
36	0.63	167.0206	[M - H] <sup>-</sup>	C <sub>5</sub> H <sub>7</sub> N <sub>4</sub> O <sub>3</sub>	167.0205	0.1	<b>uric acid</b>	37–39	G
71	1.95	203.0817	[M - H] <sup>-</sup>	C <sub>11</sub> H <sub>12</sub> N <sub>2</sub> O <sub>2</sub>	203.0821	0.4	<b>tryptophan</b>	34,56	G
384	11.70	508.3403	[M - CH <sub>3</sub> ] <sup>-</sup>	C <sub>26</sub> H <sub>34</sub> NO <sub>7</sub> P	508.3403	0.0	<b>lysoPC(18:0/0:0)</b>	19,57	G
84	8.41	223.1331	[M - H] <sup>-</sup>	C <sub>13</sub> H <sub>20</sub> O <sub>3</sub>	223.1334	0.3	13-oxo-9,11-tridecadienoic acid	58	G
157	7.06	273.1703	[M - H] <sup>-</sup>	C <sub>14</sub> H <sub>26</sub> O <sub>5</sub>	273.1702	0.1	3-hydroxytetradecanedioic acid	59,60	G
176	7.61	287.1854	[M - H] <sup>-</sup>	C <sub>15</sub> H <sub>28</sub> O <sub>5</sub>	287.1858	0.4	6-hydroxypentadecanedioic acid		G
55	5.21	185.0812	[M - H] <sup>-</sup>	C <sub>9</sub> H <sub>14</sub> O <sub>4</sub>	185.0814	0.2	5-(2-methylpropyl)-2-oxooxolane-3-carboxylic acid	61	G
							5-butyl-2-oxooxolane-3-carboxylic acid		
343	9.77	476.2772	[M - H] <sup>-</sup>	C <sub>23</sub> H <sub>44</sub> NO <sub>7</sub> P	476.2777	0.5	lysoPE(0:0/18:2)	62	G
							lysoPE(18:2/0:0)		
429	9.80	578.3450	[M + CH <sub>3</sub> COO] <sup>-</sup>	C <sub>26</sub> H <sub>50</sub> NO <sub>7</sub> P	578.3458	0.8	lysoPC(18:2/0:0)	19	G
409	5.46	541.2639	[M - H] <sup>-</sup>	C <sub>27</sub> H <sub>42</sub> O <sub>11</sub>	541.2649	1.0	cortolone-3-glucuronide	63,64	G
386	6.92	511.2900	[M - H] <sup>-</sup>	C <sub>27</sub> H <sub>44</sub> O <sub>9</sub>	511.2907	0.7	pregnanetriol glucuronide		G
173	8.19	285.1920	[M - H] <sup>-</sup>	C <sub>19</sub> H <sub>26</sub> O <sub>2</sub>	285.1855	6.5	androstenedione	65	G
393	8.12	517.3015	-	-	-	-	-		G
438	7.04	600.2572	-	-	-	-	-		G
147	0.55	266.8028	-	-	-	-	-		G
82	8.12	215.1281	-	-	-	-	-		G
43	9.56	171.1383	[M - H] <sup>-</sup>	C <sub>10</sub> H <sub>20</sub> O <sub>2</sub>	171.1385	0.2	<b>decanoic acid (capric acid)</b>		F
223	6.77	331.1753	[M - H] <sup>-</sup>	C <sub>16</sub> H <sub>28</sub> O <sub>7</sub>	331.1757	0.4	menthol glucuronide	66–69	F
							citronellol glucuronide		
63	7.19	195.1020	[M - H] <sup>-</sup>	C <sub>10</sub> H <sub>16</sub> N <sub>2</sub> O <sub>2</sub>	195.1134	11.4	l- $\alpha$ -amino-1H-pyrrole-1-hexanoic acid		F
376	9.63	504.3081	[M - CH <sub>3</sub> ] <sup>-</sup>	C <sub>26</sub> H <sub>50</sub> NO <sub>7</sub> P	504.309	0.9	lysoPC(0:0/18:2) <sup>b</sup>	62	F
412	8.86	545.3323	-	-	-	-	-		F
211	4.06	311.1387	[M - H] <sup>-</sup>	C <sub>18</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub>	311.1396	0.9	<b>phenylalanyl phenylalanine</b>	35	E
250	5.70	383.1521	[M - H] <sup>-</sup>	C <sub>19</sub> H <sub>28</sub> O <sub>6</sub> S	383.1528	0.7	3 $\beta$ ,16 $\alpha$ -dihydroxyandrostene sulfate		E
107	5.40	245.0480	[M - H] <sup>-</sup>	C <sub>10</sub> H <sub>14</sub> O <sub>5</sub> S	245.0484	0.4	2- <i>tert</i> -butyl-1,4-benzenediol sulfate	70	E
76	2.64	212.0016	[M - H] <sup>-</sup>	C <sub>8</sub> H <sub>7</sub> NO <sub>4</sub> S	212.0018	0.2	<b>indoxyl sulfuric acid</b>	12,42	D
214	9.87	311.2211	[M - H] <sup>-</sup>	C <sub>18</sub> H <sub>32</sub> O <sub>4</sub>	311.2222	1.1	9,10-dihydroxy-12Z,15Z-octadecadienoic acid (9,10-DiHODE)	71,72	D
							12,13-dihydroxy-9Z,15Z-octadecadienoic acid (12,13-DiHODE)		
							15,16-dihydroxy-9Z,12Z-octadecadienoic acid (15,16-DiHODE)		
444	6.82	613.3583	[M - H] <sup>-</sup>	C <sub>32</sub> H <sub>54</sub> O <sub>11</sub>	613.3588	0.5	27-nor-5 $\beta$ -cholestane-3 $\alpha$ ,7 $\alpha$ ,12 $\alpha$ ,24,25-pentol glucuronide	73,74	D
174	9.35	285.2059	[M - H] <sup>-</sup>	C <sub>16</sub> H <sub>30</sub> O <sub>4</sub>	285.2066	0.7	<b>hexadecanedioic acid</b>	75	C
128	2.69	263.1023	[M - H] <sup>-</sup>	C <sub>13</sub> H <sub>16</sub> N <sub>2</sub> O <sub>4</sub>	263.1032	0.9	<b>phenylacetylglutamine</b>	76,77	C
153	14.80	269.2475	[M - H] <sup>-</sup>	C <sub>17</sub> H <sub>34</sub> O <sub>2</sub>	269.2481	0.6	<b>heptadecanoic acid</b>	57	C
398	7.06	528.2630	[M - H] <sup>-</sup>	C <sub>26</sub> H <sub>43</sub> NO <sub>8</sub> S	528.2631	0.1	n-[(3 $\alpha$ ,5 $\beta$ ,7 $\beta$ )-7-hydroxy-24-oxo-3-(sulfoxy)cholan-24-yl]-glycine	78	C
							n-[(3 $\alpha$ ,5 $\beta$ ,7 $\alpha$ )-3-hydroxy-24-oxo-7-(sulfoxy)cholan-24-yl]-glycine		
							glycochenodeoxycholate-3-sulfate		
93	6.36	229.0534	[M - H] <sup>-</sup>	C <sub>10</sub> H <sub>14</sub> O <sub>4</sub> S	229.0535	0.1	5-isopropyl-2-methylphenol sulfate (carvacrol sulfate)	79	C
360	8.16	489.2692	-	-	-	-	-		C
448	8.51	621.3273	-	-	-	-	-		C
364	5.57	495.2228	[M - H] <sup>-</sup>	C <sub>25</sub> H <sub>36</sub> O <sub>10</sub>	495.2230	0.2	5'-carboxy- $\alpha$ -chromanol glucuronide	55	B
21	5.16	144.0471	[M - H] <sup>-</sup>	C <sub>9</sub> H <sub>7</sub> NO	144.0449	2.2	indole-3-carboxaldehyde	53,54	B
404	7.28	537.2501	-	-	-	-	-		B
242	7.66	369.1740	[M - H] <sup>-</sup>	C <sub>19</sub> H <sub>30</sub> O <sub>5</sub> S	369.1736	0.4	androsterone sulfate	43–45	A
							5 $\alpha$ -dihydrotestosterone sulfate		
							etiocholanolone sulfate		
237	11.34	365.2680	-	-	-	-	-		A

<sup>a</sup>Metabolites confirmed by retention time matching with commercially available standards are highlighted in bold font. Abbreviations: lysoPC, lysophosphatidylcholine; lysoPE, lysophosphatidylethanolamine. <sup>b</sup>Not in HMDB.

and fragmentation patterns were manually analyzed, as well (Figure 6E). Finally, standards of commercially available metabolites were subject to UPLC-MS and MS/MS to verify the identity of the candidates by retention time and mass spectral matching. Of the 40 spectral features found in panel A, 31 were identified by HRAM MS and MS/MS, with 10 further confirmed chromatographically by standards. The set of 31 metabolites provided 90.9% sensitivity, 91.3% specificity, and 91.1% accuracy, whereas the 10 differential metabolites confirmed by standards, when considered alone, provided 79.9% sensitivity, 70.6% specificity, and 76.3% accuracy (Table S2, Supporting Information). It should be noted that, among the 31 identified metabolites, 1- $\alpha$ -amino-1*H*-pyrrole-1-hexanoic acid (feature code 63) had the highest mass error (11.4 mDa), and its identity should be viewed as tentative. However, a classification model built using the set of 30 metabolites excluding feature 63 still provided 92.8% sensitivity, 89.2% specificity, and 91.2% accuracy.

### Biological Relevance of the IVDMA Metabolites

Table 2 summarizes the results from the chemical identification workflow described above for the 40 discriminant features. Those metabolites with chromatographic identity confirmation by retention time matching with standards are shown in bold and therefore can be viewed as the ones with the higher confidence in the panel. Several discriminant metabolites were identified as fatty acids, amino acids, lysophospholipids, and bile acids, suggesting alterations in their respective metabolism. Previous findings have shown abnormality in fatty acid<sup>33</sup> and amino acid<sup>12,34,35</sup> metabolism in PCa patients. Alterations in fatty acid metabolism through an enhanced  $\beta$ -oxidation pathway have been suggested to provide bioenergy for abnormal cell proliferation.<sup>33</sup> Among the different lysophospholipids identified that may play a role in cell signaling,<sup>36</sup> lysoPC(18:2) and lysoPC(18:0) have been reported as biomarkers for PCa detection within a panel of plasma lipids.<sup>19</sup> Uric acid has also been suggested to be a disease risk marker due to its pro-inflammatory properties,<sup>37,38</sup> and a prospective epidemiological study demonstrated positive association between serum uric acid levels and risk of PCa development.<sup>39</sup> In addition, elevated concentrations of serum uric acid are often found due to tumor lysis syndrome observed as a result of cancer therapy.<sup>40</sup> Interestingly, indoxyl sulfate, a toxic product of dietary tryptophan metabolism that accumulates in the blood of patients with impaired renal function,<sup>41</sup> was also identified among the 40 discriminant features. The reason behind elevated indoxyl sulfate in serum of PCa patients is not yet fully understood; nevertheless, this nephrotoxic metabolite likely contributes to the disease or its complications via multiple mechanisms, including enhanced oxidative stress due to decreased levels of glutathione.<sup>42</sup>

Perhaps the most salient finding resulting from the chemical identification workflow is that many differentiating metabolites belong to the steroid hormone biosynthesis pathway. As illustrated in Figure S4, the pathway supplies androgens<sup>43–45</sup> such as testosterone and 5 $\alpha$ -dihydrotestosterone, to support the growth of androgen-dependent PCa.<sup>46</sup> An average increase of pregnanetriol and androstenedione concentrations in PCa serum suggests that there is a metabolic alteration of the steroid pathway that mimics congenital adrenal hyperplasia (CAH), a metabolic disease that is accompanied by androgen excess due to the diversion of 17-hydroxyprogesterone into the pathway for androgen biosynthesis.<sup>47,48</sup> In addition, the average decrease

of azelaic acid concentration in the serum of PCa patients, an inhibitor of 5 $\alpha$ -reductase,<sup>49</sup> suggests the disinhibition of 5 $\alpha$ -reductase, an enzyme that catalyzes the synthesis of highly active androgen 5 $\alpha$ -dihydrotestosterone to support PCa growth. Indeed, azelaic acid, which has a large contribution in the models, has been postulated to be a potential antitumoral agent.<sup>50</sup>

Table 2 also shows the identification of several xenobiotics that can be grouped into two classes according to their origin. Menthol, citronellol, carvacrol, and *t*-butylhydroquinone are most likely related to food components. Assuming that both PCa patients and healthy individuals were equally exposed, on average, to food components/additives, their different metabolism could explain the different levels of these xenometabolites in serum. For example, the terpenoids menthol, carvacrol and citronellol are metabolized by CYP2A6,<sup>51,52</sup> which is also involved in steroid metabolism. As a result, average lower concentrations of these terpenoids relative to healthy individuals may be suggestive of higher activity of CYP2A6 in PCa patients, supporting inclusion of these xenometabolites in the models. The second group of xenobiotics comprises indole-3-carboxaldehyde and 5'-carboxy- $\alpha$ -chromanol glucuronide, which could possibly result from the consumption of dietary supplements used by cancer patients. Self-medicating with an over-the-counter indole-3-carbinol (I3C) supplement may explain the increased average concentration of indole-3-carboxaldehyde in PCa serum.<sup>53</sup> Indeed, indole-3-carboxaldehyde demonstrated activity against prostate cancer in both in vitro and in vivo models.<sup>54</sup> Similarly,  $\alpha$ -tocopherol, a form of vitamin E and a precursor of 5'-carboxy- $\alpha$ -chromanol glucuronide, has been suggested to influence the development of PCa due to its antioxidant activity.<sup>55</sup> As humans do not normally produce indole-3-carbaldehyde or 5'-carboxy- $\alpha$ -chromanol, and their consideration in the models may reflect dietary supplementation differences rather than endogenous metabolic differences, PCa detection was attempted using 28 of the 31 identified metabolites, excluding from the SVM classification model two metabolites which might result from dietary supplementation and one metabolite with highest mass error (1- $\alpha$ -amino-1*H*-pyrrole-1-hexanoic acid). This modified classification model provided 89.7% sensitivity, 90.7% specificity, and 90.2% accuracy (Table S2, Supporting Information), indicating that the three excluded metabolites had little effect on the overall assay performance, as supported by their low weights in panel A (Figure 5 and Figure S3, Supporting Information).

### CONCLUSIONS

The present study shows the combined application of UPLC-MS/MS and machine learning methods to develop a metabolite-based IVDMA that predicts the presence of PCa in serum samples with high classification sensitivity, specificity, and accuracy. A panel of 40 metabolic spectral features was found to be differential with 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. Of further significance, the detection performance of the IVDMA was proven to be higher than the prevalent PSA test, highlighting that a combination of multiple discriminant features yields higher predictive power for PCa detection than the univariate analysis of a single marker. Within the discriminant panel, 31 metabolites were identified by HRAM MS and MS/MS, with 10 further confirmed chromatographically by standards. Fatty acids, amino acids, lysophospholipids, and bile acids have been identified among the

discriminant metabolites, suggesting alterations in their metabolism. Additionally, several metabolites were mapped to the steroid hormone biosynthesis pathway. These observations demonstrate some of the plausible metabolic alterations in PCa and provide further insight into the biological pathway changes associated with the disease. The combination of multiple metabolites that yield a single, patient-specific result for disease detection is the strength of the IVDMA developed in the present work. When the assay is based on the 28 identified disease-related metabolites, PCa can still be detected with 89.7% sensitivity, 90.7% specificity, and 90.2% accuracy. If higher throughput analysis and lower analysis cost and complexity are desired, 13 metabolites that were found to be present in 90% of the entire sample cohort would still provide high classification sensitivity (88.3%), specificity (80.3%), and accuracy (85.0%) for cancerous and healthy samples. Therefore, this assay shows promise toward its implementation in the clinical laboratory setting once it is fully validated by the examination of a larger patient cohort through targeted assays.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

Figure S1: Visualization of the PCa metabolic scores obtained by SVMs model validation iterations. Figure S2: Principal component analysis of the metabolic data. Table S1: Gleason scores for PCa patients. Table S2: IVDMA performance for identified metabolites. Figure S3: SVM weights for the discriminant metabolic features from panels A–G. Figure S4: KEGG steroid hormone biosynthesis pathway (hsa00140). This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [facundo.fernandez@chemistry.gatech.edu](mailto:facundo.fernandez@chemistry.gatech.edu). Phone: 404-385-4432. Fax: 404-385-3399.

### Present Address

#Centro de Investigaciones en Bionanociencias (CIBION), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2390, C1425FQD, Ciudad de Buenos Aires, Argentina.

### Author Contributions

<sup>†</sup>These authors contributed equally to this work.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Drs. Nikhil Shah and Rajesh Laungani are acknowledged for sample collection at Saint Joseph's Hospital, Atlanta, GA (USA). F.M.F. acknowledges support from a Vasser-Wooley professorship. Funding was provided by Saint Joseph's Mercy Foundation to J.F.M.

## ■ REFERENCES

- (1) *Cancer Facts & Figures 2013*; American Cancer Society, Atlanta, GA, 2013.
- (2) *Cancer Facts & Figures 2010*; American Cancer Society, Atlanta, GA, 2010.
- (3) Nadler, R. B.; Humphrey, P. A.; Smith, D. S.; Catalona, W. J.; Ratliff, T. L. Effect of inflammation and benign prostatic hyperplasia

on elevated serum prostate-specific antigen levels. *J. Urol.* **1995**, *154* (2), 407–413.

(4) Tombal, B. Over- and underdiagnosis of prostate cancer: the dangers. *Eur. Urol. Suppl.* **2006**, *5* (6), 511–513.

(5) Bickers, B.; Aukim-Hastie, C. New molecular biomarkers for the prognosis and management of prostate cancer—the post PSA era. *Anticancer Res.* **2009**, *29* (8), 3289–3298.

(6) Heijnsdijk, E. A. M.; der Kinderen, A.; Wever, E. M.; Draisma, G.; Roobol, M. J.; de Koning, H. J. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *Br. J. Cancer* **2009**, *101* (11), 1833–1838.

(7) Draisma, G.; Etzioni, R.; Tsodikov, A.; Mariotto, A.; Wever, E.; Gulati, R.; Feuer, E.; de Koning, H. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J. Natl. Cancer Inst.* **2009**, *101* (6), 374–383.

(8) Thompson, I. M.; Pauler, D. K.; Goodman, P. J.; Tangen, C. M.; Lucia, M. S.; Parnes, H. L.; Minasian, L. M.; Ford, L. G.; Lippman, S. M.; Crawford, E. D.; Crowley, J. J.; Coltman, C. A. Prevalence of prostate cancer among men with a prostate-specific antigen level  $\leq$  4.0 ng per milliliter. *N. Engl. J. Med.* **2004**, *350* (22), 2239–2246.

(9) Trock, B. J. Application of metabolomics to prostate cancer. *Urol. Oncol.: Semin. Orig. Invest.* **2011**, *29* (5), 572–581.

(10) Chace, D. H.; Kalas, T. A.; Naylor, E. W. The application of tandem mass spectrometry to neonatal screening for inherited disorders of intermediary metabolism. *Annu. Rev. Genomics Hum. Genet.* **2002**, *3*, 17–45.

(11) Piraud, M.; Vianey-Saban, C.; Petritis, K.; Elfakir, C.; Steghens, J. P.; Morla, A.; Bouchu, D. ESI-MS/MS analysis of underivatized amino acids: a new tool for the diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionisation mode. *Rapid Commun. Mass Spectrom.* **2003**, *17* (12), 1297–1311.

(12) Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J. D.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X. H.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009**, *457* (7231), 910–914.

(13) Jentzmik, F.; Stephan, C.; Miller, K.; Schrader, M.; Erbersdobler, A.; Kristiansen, G.; Lein, M.; Jung, K. Sarcosine in urine after digital rectal examination fails as a marker in prostate cancer detection and identification of aggressive tumours. *Eur. Urol.* **2010**, *58* (1), 12–18.

(14) Schalken, J. A. Is urinary sarcosine useful to identify patients with significant prostate cancer? The trials and tribulations of biomarker development. *Eur. Urol.* **2010**, *58* (1), 19–20.

(15) Struys, E. A.; Heijboer, A. C.; van Moorselaar, J.; Jakobs, C.; Blankenstein, M. A. Serum sarcosine is not a marker for prostate cancer. *Ann. Clin. Biochem.* **2010**, *47*, 282–282.

(16) Swanson, M. G.; Vigneron, D. B.; Tabatabai, Z. L.; Males, R. G.; Schmitt, L.; Carroll, P. R.; James, J. K.; Hurd, R. E.; Kurhanewicz, J. Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted postsurgical prostate tissues. *Magn. Reson. Med.* **2003**, *50* (5), 944–954.

(17) Swanson, M. G.; Zektzer, A. S.; Tabatabai, Z. L.; Simko, J.; Jarso, S.; Keshari, K. R.; Schmitt, L.; Carroll, P. R.; Shinohara, K.; Vigneron, D. B.; Kurhanewicz, J. Quantitative analysis of prostate metabolites using H-1 HR-MAS spectroscopy. *Magn. Reson. Med.* **2006**, *55* (6), 1257–1264.

(18) Thysell, E.; Surowiec, I.; Hornberg, E.; Crnalic, S.; Widmark, A.; Johansson, A. I.; Stattin, P.; Bergh, A.; Moritz, T.; Antti, H.; Wikstro, P. Metabolomic characterization of human prostate cancer bone metastases reveals increased levels of cholesterol. *PLoS One* **2010**, *5* (12), e14175.

(19) Zhou, X. C.; Mao, J. H.; Ai, J. M.; Deng, Y. P.; Roth, M. R.; Pound, C.; Henegar, J.; Welti, R.; Bigler, S. A. Identification of plasma lipid biomarkers for prostate cancer by lipidomics and bioinformatics. *PLoS One* **2012**, *7* (11), e48889.

- (20) Boccard, J.; Veuthey, J. L.; Rudaz, S. Knowledge discovery in metabolomics: an overview of MS data handling. *J. Sep. Sci.* **2010**, *33* (3), 290–304.
- (21) Eliasson, M.; Rannar, S.; Trygg, J. From data processing to multivariate validation—essential steps in extracting interpretable information from metabolomics data. *Curr. Pharm. Biotechnol.* **2011**, *12* (7), 996–1004.
- (22) Madsen, R.; Lundstedt, T.; Trygg, J. Chemometrics in metabolomics—a review in human disease diagnosis. *Anal. Chim. Acta* **2010**, *659* (1–2), 23–33.
- (23) Guan, W.; Zhou, M.; Hampton, C.; Benigno, B.; Walker, L. D.; Gray, A.; McDonald, J.; Fernandez, F. Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinf.* **2009**, *10* (1), 259.
- (24) Zhou, M. S.; Guan, W.; Walker, L. D.; Mezencev, R.; Benigno, B. B.; Gray, A.; Fernandez, F. M.; McDonald, J. F. Rapid mass spectrometric metabolic profiling of blood sera detects ovarian cancer with high accuracy. *Cancer Epidemiol., Biomarkers Prev.* **2010**, *19* (9), 2262–2271.
- (25) Likhov, P. G.; Kharybin, O. N.; Archakov, A. I. Diagnosis of lung cancer based on direct-infusion electrospray mass spectrometry of blood plasma metabolites. *Int. J. Mass Spectrom.* **2012**, *309*, 200–205.
- (26) Frickenschmidt, A.; Frohlich, H.; Bullinger, D.; Zell, A.; Laufer, S.; Gleiter, C. H.; Liebich, H.; Kammerer, B. Metabonomics in cancer diagnosis: mass spectrometry-based profiling of urinary nucleosides from breast cancer patients. *Biomarkers* **2008**, *13* (4), 435–449.
- (27) Zhang, Z. An in vitro diagnostic multivariate index assay (IVDMIA) for ovarian cancer: harvesting the power of multiple biomarkers. *Rev. Obstet. Gynecol.* **2012**, *5* (1), 35–41.
- (28) Brereton, R. G.; Lloyd, G. R. Support vector machines for classification and regression. *Analyst* **2010**, *135* (2), 230–267.
- (29) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. METLIN: a metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27* (6), 747–751.
- (30) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorn Dahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. HMDB 3.0: the human metabolome database in 2013. *Nucleic Acids Res.* **2013**, *41* (D1), D801–D807.
- (31) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45* (7), 703–714.
- (32) Nicholson, J. K.; Lindon, J. C. Systems biology: metabonomics. *Nature* **2008**, *455* (7216), 1054–1056.
- (33) Liu, Y. Fatty acid oxidation is a dominant bioenergetic pathway in prostate cancer. *Prostate Cancer Prostatic Dis.* **2006**, *9* (3), 230–234.
- (34) Ebenezar, J.; Pu, Y.; Wang, W. B.; Liu, C. H.; Alfano, R. R. Stokes shift spectroscopy pilot study for cancerous and normal prostate tissues. *Appl. Opt.* **2012**, *51* (16), 3642–3649.
- (35) Fu, Y. M.; Lin, H.; Liu, X.; Fang, W.; Meadows, G. G. Cell death of prostate cancer cells by specific amino acid restriction depends on alterations of glucose metabolism. *J. Cell Physiol.* **2010**, *224* (2), 491–500.
- (36) Peyruchaud, O. Novel implications for lysophospholipids, lysophosphatidic acid and sphingosine 1-phosphate, as drug targets in cancer. *Anti-Cancer Agents Med. Chem.* **2009**, *9* (4), 381–391.
- (37) Brys, M.; Morel, A.; Forma, E.; Krzeslak, A.; Wilkosz, J.; Rozanski, W.; Olas, B. Relationship of urinary isoprostanes to prostate cancer occurrence. *Mol. Cell. Biochem.* **2013**, *372* (1–2), 149–153.
- (38) Fini, M.; Elias, A.; Johnson, R.; Wright, R. Contribution of uric acid to cancer risk, recurrence, and mortality. *Clin. Transl. Med.* **2012**, *1* (1), 16.
- (39) Kolonel, L.; Yoshizawa, C.; Nomura, A.; Stemmermann, G. Relationship of serum uric acid to cancer occurrence in a prospective male cohort. *Cancer Epidemiol. Biomarkers Prev.* **1994**, *3* (3), 225–228.
- (40) Tsimberidou, A.; Keating, M. Hyperuricemic syndromes in cancer patients. *Contrib. Nephrol.* **2005**, *147*, 47–60.
- (41) Barreto, F. C.; Barreto, D. V.; Liabeuf, S.; Meert, N.; Glorieux, G.; Temmar, M.; Choukroun, G.; Vanholder, R.; Massy, Z. A. European uremic toxin work grp, e.u.t., serum indoxyl sulfate is associated with vascular disease and mortality in chronic kidney disease patients. *Clin. J. Am. Soc. Nephrol.* **2009**, *4* (10), 1551–1558.
- (42) Dou, L.; Jourde-Chiche, N.; Faure, V.; Cerini, C.; Berland, Y.; Dignat-George, F.; Brunet, P. The uremic solute indoxyl sulfate induces oxidative stress in endothelial cells. *J. Thromb. Haemost.* **2007**, *5* (6), 1302–1308.
- (43) Kliman, B.; Prout, G. R.; Maclaughlin, R. A.; Daly, J. J.; Griffin, P. P. Altered androgen metabolism in metastatic prostate cancer. *J. Urol.* **1978**, *119* (5), 623–626.
- (44) Titus, M. A.; Schell, M. J.; Lih, F. B.; Tomer, K. B.; Mohler, J. L. Testosterone and dihydrotestosterone tissue levels in recurrent prostate cancer. *Clin. Cancer Res.* **2005**, *11* (13), 4653–4657.
- (45) Zumoff, B.; Levin, J.; Strain, G. W.; Rosenfeld, R. S.; O'Connor, J.; Freed, S. Z.; Kream, J.; Whitmore, W. S.; Fukushima, D. K.; Hellman, L. Abnormal levels of plasma hormones in men with prostate-cancer: evidence toward a 2-disease theory. *Prostate* **1982**, *3* (6), 579–588.
- (46) Knudsen, K. E.; Scher, H. I. Starving the addiction: new opportunities for durable suppression of AR signaling in prostate cancer. *Clin. Cancer Res.* **2009**, *15* (15), 4792–4798.
- (47) Horton, R.; Frasier, S. D. Androstenedione and its conversion to plasma testosterone in congenital adrenal hyperplasia. *J. Clin. Invest.* **1967**, *46* (6), 1003.
- (48) Speiser, P. W.; White, P. C. Congenital adrenal hyperplasia. *N. Engl. J. Med.* **2003**, *349* (8), 776–788.
- (49) Passi, S.; Picardo, M.; De Luca, C.; Nazzaro-Porro, M. Mechanism of azelaic acid action in acne. *G. Ital. Dermatol. Venereol.* **1989**, *124* (10), 455–463.
- (50) Breathnach, A. S. Azelaic acid: potential as a general antitumoural agent. *Med. Hypotheses* **1999**, *52* (3), 221–226.
- (51) Dong, R. H.; Fang, Z. Z.; Zhu, L. L.; Ge, G. B.; Cao, Y. F.; Li, X. B.; Hu, C. M.; Yang, L.; Liu, Z. Y. Identification of CYP isoforms involved in the metabolism of thymol and carvacrol in human liver microsomes (HLMs). *Pharmazie* **2012**, *67* (12), 1002–1006.
- (52) Miyazawa, M.; Marumoto, S.; Takahashi, T.; Nakahashi, H.; Haigou, R.; Nakanishi, K. metabolism of (+)- and (–)-menthols by CYP2A6 in human liver microsomes. *J. Oleo Sci.* **2011**, *60* (3), 127–132.
- (53) Anderton, M. J.; Manson, M. M.; Verschoyle, R. D.; Gescher, A.; Lamb, J. H.; Farmer, P. B.; Steward, W. P.; Williams, M. L. Pharmacokinetics and tissue disposition of indole-3-carbinol and its acid condensation products after oral administration to mice. *Clin. Cancer Res.* **2004**, *10* (15), 5233–5241.
- (54) Souli, E.; Machluf, M.; Morgenstern, A.; Sabo, E.; Yannai, S. Indole-3-carbinol (I3C) exhibits inhibitory and preventive effects on prostate tumors in mice. *Food Chem. Toxicol.* **2008**, *46* (3), 863–870.
- (55) Heinonen, O. P.; Albanes, D.; Virtamo, J.; Taylor, P. R.; Huttunen, J. K.; Hartman, A. M.; Haapakoski, J.; Malila, N.; Rautalahti, M.; Ripatti, S.; Maenpaa, H.; Teerenhovi, L.; Koss, L.; Virolainen, M.; Edwards, B. K. Prostate cancer and supplementation with  $\alpha$ -tocopherol and  $\beta$ -carotene: incidence and mortality in a controlled trial. *J. Natl. Cancer Inst.* **1998**, *90* (6), 440–446.
- (56) Prendergast, G. C. Cancer: why tumours eat tryptophan. *Nature* **2011**, *478* (7368), 192–194.
- (57) Crowe, F. L.; Allen, N. E.; Appleby, P. N.; Overvad, K.; Aardstrup, I. V.; Johnsen, N. F.; Tjonneland, A.; Linseisen, J.; Kaaks, R.; Boeing, H.; Kroger, J.; Trichopoulou, A.; Zavitsanou, A.; Trichopoulos, D.; Sacerdote, C.; Palli, D.; Tumino, R.; Agnoli, C.; Kiemene, L. A.; Bueno-de-Mesquita, H. B.; Chirlaque, M. D.; Ardanaz, E.; Larranaga, N.; Quiros, J. R.; Sanchez, M. J.; Gonzalez, C. A.; Stattin, P.; Hallmans, G.; Bingham, S.; Khaw, K. T.; Rinaldi, S.;

Slimani, N.; Jenab, M.; Riboli, E.; Key, T. J. Fatty acid composition of plasma phospholipids and risk of prostate cancer in a case-control analysis nested within the European Prospective Investigation into Cancer and Nutrition. *Am. J. Clin. Nutr.* **2008**, *88* (5), 1353–1363.

(58) Glasgow, W.; Eling, T. Structural requirements for enhancement of EGF-dependent DNA synthesis by oxygenated metabolites of linoleic acid. In *Eicosanoids and Other Bioactive Lipids in Cancer, Inflammation, and Radiation Injury 2*; Honn, K., Nigam, S., Marnett, L., Eds.; Springer: New York, 1997; Vol. 400, pp 507–512.

(59) Tserng, K. Y.; Jin, S. J. Metabolic origin of urinary 3-hydroxy dicarboxylic acids. *Biochemistry* **1991**, *30* (9), 2508–2514.

(60) Zha, S.; Ferdinandusse, S.; Hicks, J. L.; Denis, S.; Dunn, T. A.; Wanders, R. J.; Luo, J.; De Marzo, A. M.; Isaacs, W. B. Peroxisomal branched chain fatty acid  $\beta$ -oxidation pathway is upregulated in prostate cancer. *Prostate* **2005**, *63* (4), 316–323.

(61) Pettersson, A.; Kasperzyk, J. L.; Kenfield, S. A.; Richman, E. L.; Chan, J. M.; Willett, W. C.; Stampfer, M. J.; Mucci, L. A.; Giovannucci, E. L. Milk and dairy consumption among men with prostate cancer and risk of metastases and prostate cancer death. *Cancer Epidemiol., Biomarkers Prev.* **2012**, *21* (3), 428–436.

(62) Sun, H.; Zhang, A. H.; Yan, G. L.; Piao, C. Y.; Li, W. Y.; Sun, C.; Wu, X. H.; Li, X. H.; Chen, Y.; Wang, X. J. Metabolomic analysis of key regulatory metabolites in hepatitis C virus-infected tree shrews. *Mol. Cell. Proteomics* **2013**, *12* (3), 710–719.

(63) Cho, H. J.; Kim, J. D.; Lee, W. Y.; Chung, B. C.; Choi, M. H. Quantitative metabolic profiling of 21 endogenous corticosteroids in urine by liquid chromatography-triple quadrupole-mass spectrometry. *Anal. Chim. Acta* **2009**, *632* (1), 101–108.

(64) Schatzl, G.; Reiter, W. J.; Thurnridl, T.; Waldmuller, J.; Roden, M.; Soregi, S.; Madersbacher, S. Endocrine patterns in patients with benign and malignant prostatic diseases. *Prostate* **2000**, *44* (3), 219–224.

(65) Barrett-Connor, E.; Garland, C.; McPhillips, J. B.; Khaw, K. T.; Wingard, D. L. A prospective, population-based study of androstenedione, estrogens, and prostatic cancer. *Cancer Res.* **1990**, *50* (1), 169–173.

(66) Li, Q.; Wang, X.; Yang, Z.; Wang, B.; Li, S. Menthol induces cell death via the TRPM8 channel in the human bladder cancer cell line T24. *Oncology* **2009**, *77* (6), 335–341.

(67) Lin, J. P.; Lu, H. F.; Lee, J. H.; Lin, J. G.; Hsia, T. C.; Wu, L. T.; Chung, J. G. (–)-Menthol inhibits DNA topoisomerases I, II  $\alpha$  and  $\beta$  and promotes NF- $\kappa$ B expression in human gastric cancer SNU-5 cells. *Anticancer Res.* **2005**, *25* (3B), 2069–2074.

(68) Sacchetti, G.; Maietti, S.; Muzzoli, M.; Scaglianti, M.; Manfredini, S.; Radice, M.; Bruni, R. Comparative evaluation of 11 essential oils of different origin as functional antioxidants, antiradicals and antimicrobials in foods. *Food Chem.* **2005**, *91* (4), 621–632.

(69) Brito, R. G.; Guimaraes, A. G.; Quintans, J. S. S.; Santos, M. R. V.; De Sousa, D. P.; Badaue-Passos, D., Jr.; de Lucca, W., Jr.; Brito, F. A.; Barreto, E. O.; Oliveira, A. P.; Quintans, L. J., Jr. Citronellol, a monoterpene alcohol, reduces nociceptive and inflammatory activities in rodents. *J. Nat. Med.* **2012**, *66* (4), 637–644.

(70) Lau, A.; Villeneuve, N. F.; Sun, Z.; Wong, P. K.; Zhang, D. D. Dual roles of Nrf2 in cancer. *Pharmacol. Res.* **2008**, *58* (5–6), 262–270.

(71) Collett, G. P.; Betts, A. M.; Johnson, M. I.; Pulimood, A. B.; Cook, S.; Neal, D. E.; Robson, C. N. Peroxisome proliferator-activated receptor  $\alpha$  is an androgen-responsive gene in human prostate and is highly expressed in prostatic adenocarcinoma. *Clin. Cancer Res.* **2000**, *6* (8), 3241–3248.

(72) Shearer, G. C.; Harris, W. S.; Pedersen, T. L.; Newman, J. W. Detection of omega-3 oxylipins in human plasma and response to treatment with omega-3 acid ethyl esters. *J. Lipid Res.* **2010**, *51* (8), 2074–2081.

(73) Karlaganis, G.; Bremmelgaard, A.; Karlaganis, V.; Sjovall, J. Precursor of 27-nor-5 $\beta$ -cholestane-3 $\alpha$ ,7 $\alpha$ ,12 $\alpha$ ,24,25-pentol in man. *J. Steroid Biochem. Mol. Biol.* **1983**, *18* (6), 725–729.

(74) Krycer, J. R.; Brown, A. J. Cholesterol accumulation in prostate cancer: A classic observation from a modern perspective. *Biochim. Biophys. Acta, Rev. Cancer* **2013**, *1835* (2), 219–229.

(75) Li, F.; Qin, X.; Chen, H.; Qiu, L.; Guo, Y.; Liu, H.; Chen, G.; Song, G.; Wang, X.; Li, F.; Guo, S.; Wang, B.; Li, Z. Lipid profiling for early diagnosis and progression of colorectal cancer using direct-infusion electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun. Mass Spectrom.* **2013**, *27* (1), 24–34.

(76) Onishi, T.; Yamakawa, K.; Franco, O. E.; Suzuki, R.; Kawamura, J. p27(Kip1) is the key mediator of phenylacetate induced cell cycle arrest in human prostate cancer cells. *Anticancer Res.* **2000**, *20* (5A), 3075–3081.

(77) Shibahara, T.; Onishi, T.; Franco, O. E.; Arima, K.; Sugimura, Y. Down-regulation of Skp2 is correlated with p27-associated cell cycle arrest induced by phenylacetate in human prostate cancer cells. *Anticancer Res.* **2005**, *25* (3B), 1881–1888.

(78) Goldberg, A. A.; Titorenko, V. I.; Beach, A.; Sanderson, J. T. Bile acids induce apoptosis selectively in androgen-dependent and -independent prostate cancer cells. *PeerJ* **2013**, *1*, e122.

(79) Patel, B.; Shah, V. R.; Bavadekar, S. A. Anti-proliferative effects of carvacrol on human prostate cancer cell line, LNCaP. *FASEB J.* **2012**, *26*, 1037.5.