Am J Cancer Res 2024;14(1):192-209
www.ajcr.us /ISSN:2156-6976/ajcr0153576

Original Article

Urinary volatile organic compounds in prostate cancer biopsy pathologic risk stratification using logistic regression and multivariate analysis models

Sabur Badmos[1], Elizabeth Noriega-Landa[1], Kiana L Holbrook[1], George E Quaye[2], Xiaogang Su[2], Qin Gao[1,3], Angelica A Chacon[1], Eric Adams[4], Thomas J Polascik[4], Adam S Feldman[5], Michael M Annabi[6], Wen-Yee Lee[1]

[1]Department of Chemistry and Biochemistry, University of Texas at El Paso, El Paso, Texas, USA; [2]Department of Mathematical Sciences, University of Texas at El Paso, El Paso, Texas, USA; [3]PDM Biologics Analytical Operations, Gilead Sciences Inc., Oceanside, California, USA; [4]Department of Urological Surgery, Duke University Medical Center, Durham, North Carolina, USA; [5]Department of Urology, Massachusetts General Hospital, Boston, Massachusetts, USA; [6]Internal Medicine Clinic, El Paso, Texas, USA

Received September 22, 2023; Accepted December 15, 2023; Epub January 15, 2024; Published January 30, 2024

Abstract: Prostate cancer (PCa) is the second leading cause of cancer-related death in American men after lung cancer. The current PCa diagnostic method, the serum prostate-specific antigen (PSA) test, is not specific, thus, alternatives are needed to avoid unnecessary biopsies and over-diagnosis of clinically insignificant PCa. To explore the application of metabolomics in such effort, urine samples were collected from 386 male adults aged 44-93 years, including 247 patients with biopsy-proven PCa and 139 with biopsy-proven negative results. The PCa-positive group was further subdivided into two groups: low-grade (ISUP Grade Group = 1; n = 139) and intermediate/high-grade (ISUP Grade Group ≥ 2; n = 108). Volatile organic compounds (VOCs) in urine were extracted by stir bar sorptive extraction (SBSE) and analyzed using thermal desorption with gas chromatography and mass spectrometry (GC-MS). We used machine learning tools to develop and evaluate models for PCa diagnosis and prognosis. In total, 22,538 VOCs were identified in the urine samples. With regularized logistic regression, our model for PCa diagnosis yielded an area under the curve (AUC) of 0.99 and 0.88 for the training and testing sets respectively. Furthermore, the model for differentiating between low-grade and intermediate/high-grade PCa yielded an average AUC of 0.78 based on a repeated test-sample approach for cross-validation. These novel methods using urinary VOCs and logistic regression were developed to fill gaps in PCa screening and assessment of PCa grades prior to biopsy. Our study findings provide a promising alternative or adjunct to current PCa screening and diagnostic methods to better target patients for biopsy and mitigate the challenges associated with over-diagnosis and over-treatment of PCa.

Keywords: Prostate cancer, diagnostic model, VOCs, GC-MS, stir-bar supportive extraction, urinary biomarkers, risk stratification, chemometrics, logistic regression, multivariate analysis

Introduction

Prostate cancer (PCa) is a complex and heterogeneous disease that varies from small indolent low-grade tumors to large aggressive life-threatening tumors [1, 2]. Currently, PCa is the most common non-cutaneous cancer and has the second highest cancer-related death rate after lung cancer among males in the United States, accounting for nearly 11% of all cancer-related deaths caused by cancer in men [1, 3]. Disease recurrence may result in metastasis, with an average survival rate of approximately 3-5 years after diagnosis. All men are at risk of developing prostate cancer, and age is the strongest predictive risk factor, with the risk increasing significantly with age. Additionally, risk factors, including family history, race, germline mutations, diet, physical activities, and exposure to chemicals, also contribute to the development of the disease [4, 5].

Prostate cancer is predominantly detected based on elevated or rising levels of serum prostate-specific antigen (PSA), a glycoprotein primarily expressed in prostate tissue [6], fol-

lowed by multiparametric magnetic resonance imaging (mpMRI) and/or tissue prostate biopsy, which is the standard of care to verify the presence of cancer. However, elevated PSA levels are not specific to PCa, as PSA can also be elevated due to prostate infection, inflammation, benign prostatic hyperplasia (BPH), or recent ejaculation. Thus, PSA is not a reliable, adequately sensitive, or specific biomarker for prostate cancer screening. In view of this, PSA values derived from PSA screening have remained controversial due to concerns about the risk of overdiagnosis and the resulting overtreatment of low-risk PCa, which may overshadow the mortality reduction resulting from screening for high-risk PCa [7].

The Gleason grading system refers to how abnormal prostate cancer cells look and how likely it would probably advance and spread over time. A lower Gleason grade means that the cancer is slower growing and not aggressive, whereas higher numbers indicate faster-growing cancer that is more likely to spread. In the current PCa Gleason Grade Group categorization, Grade Group X (GGX) had a Gleason score lower than 6 (no cancer), while Grade Group 1 (GG1) had a Gleason score of 6. Grade Groups 2 (GG2) and 3 (GG3) had a Gleason score of 7 (3 + 4) or (4 + 3), respectively. Furthermore, GG4 has a Gleason score of 8, while GG5 has a Gleason score of 9 or 10 [8]. Patients with low-grade PCa are typically low-risk, and most patients diagnosed with low-risk PCa are recommended for active surveillance or observation instead of active PCa treatment. However, some patients with low-risk PCa still decide to pursue active treatment instead of surveillance because they have too much anxiety about the diagnosis of untreated "cancer" and/or because they do not want to continue repeating invasive testing as active surveillance, which is a cause of concern for overtreatment. Therefore, a non-invasive method of identifying patients with intermediate/high-grade PCa would be of great clinical value to help decrease the unintended harm associated with over-treatment and over-diagnosis of low-risk PCa. Furthermore, a non-invasive method of differentiating between low-grade PCa and intermediate/high-grade PCa could be of great clinical value for patients with low-grade PCa undergoing active surveillance to optimize their further work-up and intervention to minimize unnecessary invasive testing and treatment.

The field of metabolomics has emerged as an evidence-based method for early disease diagnosis, therapy monitoring, and understanding the pathogenesis of many diseases. Metabolomics studies small biomolecules, i.e., metabolites, in urine, blood, tissue, and other body fluids in biological systems. It combines analytical technique and multivariate data analysis to identify and quantify significant metabolites that are involved in different metabolic pathways in living organisms. The resulting data provide insights into a biological system's current biochemical state under the influence of intrinsic and extrinsic factors such as disease, dietary and lifestyle choices, drugs, chemical exposures, and hazardous agents. Mass spectrometry-based metabolomics techniques are the most sensitive for the simultaneous analysis of a large number of compounds [9, 10]. As the metabolites are considered a read-out of the physiology and biochemical activity of cells, metabolomics provides a valuable platform that can be explored in the identification of cancer biomarkers and tumorigenesis drivers in the individual system [11]. This method is a rapidly expanding approach in translational research and clinical diagnostics. Furthermore, metabolomics uses a bottom-up approach involving the comprehensive detection and quantitative analysis of all biological system metabolites [12-15]. Changes in metabolite concentrations, such as volatile organic compounds (VOCs), in biological fluids are frequently indicative of alterations in individuals' physiological state, making them valuable markers in diagnosing pathological conditions. The non-invasive procedures of metabolomics for use in disease detection and monitoring have become an attractive focus in recent research because of the possibility of providing rapid, cost-effective, convenient, and efficient methods for disease diagnosis and monitoring [16-18].

In a recent discovery, a Belgian Malinois shepherd dog was trained to smell odor emanating from the urine of men for PCa detection. The study revealed that trained dogs could discriminate between urine samples from men with PCa and healthy controls, and achieved a sensitivity and specificity test assessment of 91% [19]. The working principle behind these findings was that VOCs present in the urine were detected by the dogs in the urine odor, which could readily be characterized with the aid of gas chromatography-mass spectrometry (GC-

MS) or gas sensors [20-22]. In our previous study, the performance of VOCs was demonstrated to have the potential to accurately distinguish between patients with PCa and PCa negative [23]. However, there is a need to further focus on the identification of patients with intermediate- or high-grade PCa (ISUP Grade Group ≥ 2). This could help decrease the risk of over-diagnosis and over-treatment of patients with low-grade PCa (ISUP Grade Group = 1). Additionally, differentiating between low-grade PCa and intermediate/high-grade PCa could be used for disease monitoring of patients with low-risk PCa on active surveillance to determine the optimal timing for further work-up and intervention. We hereby present the first study showing the application of VOCs for the identification of patients with PCa, as well as the differentiation between patients with low-grade PCa and intermediate/high-grade PCa.

Material and methods

Chemicals and materials

All chemicals were of analytical grade or higher. Methanol (LC-MS grade) purchased from Burdick & Jackson (Muskegon, Michigan, USA) was used to prepare Mirex (internal standard, 99.0% purity; Dr. Ehrenstorfer GmbH, Augsburg, Germany) in a 100 mg/L solution. Hydrochloric acid (37%) was purchased from Sigma-Aldrich (St. Louis, Missouri, USA). Ultra-pure deionized water (DI, Milli-Q benchtop Lab water purification system, Millipore, Bedford, Massachusetts, USA) was used to prepare the 2 M HCl solution and urine samples. Stir bar sorptive extraction (SBSE) stir bars (Twister®) coated with polydimethylsiloxane with a film thickness of 10 mm × 1 mm and thermal desorption tubes (TDT) were purchased from GERSTEL (Mülheim, Germany). Perfluorotributylamine (PFTBA) was purchased from Agilent Technologies (CA, USA) for the mass spectrometer calibration.

Urine samples collection

Ethical approval and sample collection: Internal Review Board (IRB) approval (University of Texas at El Paso (UTEP) IRB 836503-9) for the study was obtained prior to the study. De-identified urine specimens were obtained from patients present at the Duke University Medical Center, Durham, North Carolina; Eastern Virginia Medical Center, Norfolk, Virginia; Michael H Annabi Internal Medicine Clinic, El Paso, Texas, and Massachusetts General Hospital, Boston, Massachusetts. Only de-identified information (such as age, race, and pathology outcomes) will be used to indicate that the samples came from a participant positive or negative for PCa.

The subjects used in this study were 386 male adults aged 44-93 years, and from this, 139 of them were PCa-negative while 247 were PCa-positive. Furthermore, the subjects were classified based on race, with African-Americans (70), Caucasians (167), Hispanics (142), and Others (7); details are provided in **Table 1**. De-identified specimens were originally obtained using Internal Review Board (IRB) approved protocols from patients prior to undergoing transrectal ultrasound-guided prostate biopsy for evaluation of elevated PSA or abnormal digital rectal examination (DRE). Based on the biopsy results, patients with a positive PCa diagnosis represented the "PCa positive" cases, while those with no evidence of PCa on biopsy were included in the "PCa negative" group. Patient urine samples and their respective demographic data were collected.

Inclusion and exclusion criteria: Patients who did not wish to participate in this study or whose urinalysis was suspicious of infection were excluded. Urine dipstick analysis was performed on all patients recruited for this study to rule out any infection before office-based transrectal ultrasound-guided biopsy. The patients' urine samples (5 mL) were collected and stored at -80°C before the urinary VOC analysis.

In this study, sex is not considered as a biological variable because PCa is a male-specific cancer. A total of 386 urine were initially sorted into two cohorts based on their biopsy results: PCa-positive (n = 247) and PCa-negative (n = 139). Patients with biopsy-proven PCa were further divided into low-grade PCa (ISUP Gleason Grade Group 1; n = 139) and intermediate/high-grade PCa (ISUP Gleason Grade Groups 2-5; n = 108). Clinically, the low-grade group is typically considered indolent PCa, while the intermediate/high-grade group is deemed to be clinically significant. Demographic data

Table 1. Age and racial distributions of patients of PCa biopsy-proven positive (low-grade and intermediate/high-grade (I/H Grade)) and PCa biopsy-proven negative subjects

A	Summary			African-Americans			Caucasians			Hispanics			Others		
Age Bracket (Years)	Total	PCa Negative	PCa Positive	Negative	Low-Grade	I/H Grade	Negative	Low-Grade	I/H Grade	Negative	Low-Grade	I/H Grade	Negative	Low-Grade	I/H Grade
44-49	10	1	9	1	2	2	0	4	1	0	0	0	0	0	0
50-59	90	28	62	6	9	8	12	23	19	9	2	0	1	0	1
60-69	146	46	100	7	13	8	12	35	27	26	10	4	1	0	3
70-79	106	47	59	3	4	5	8	11	9	35	15	15	1	0	0
80-89	29	14	15	1	1	0	2	0	1	11	9	4	0	0	0
90-99	5	3	2	0	0	0	3	0	0	0	1	1	0	0	0
Total	386	139	247	18	29	23	37	73	57	81	37	24	3	0	4

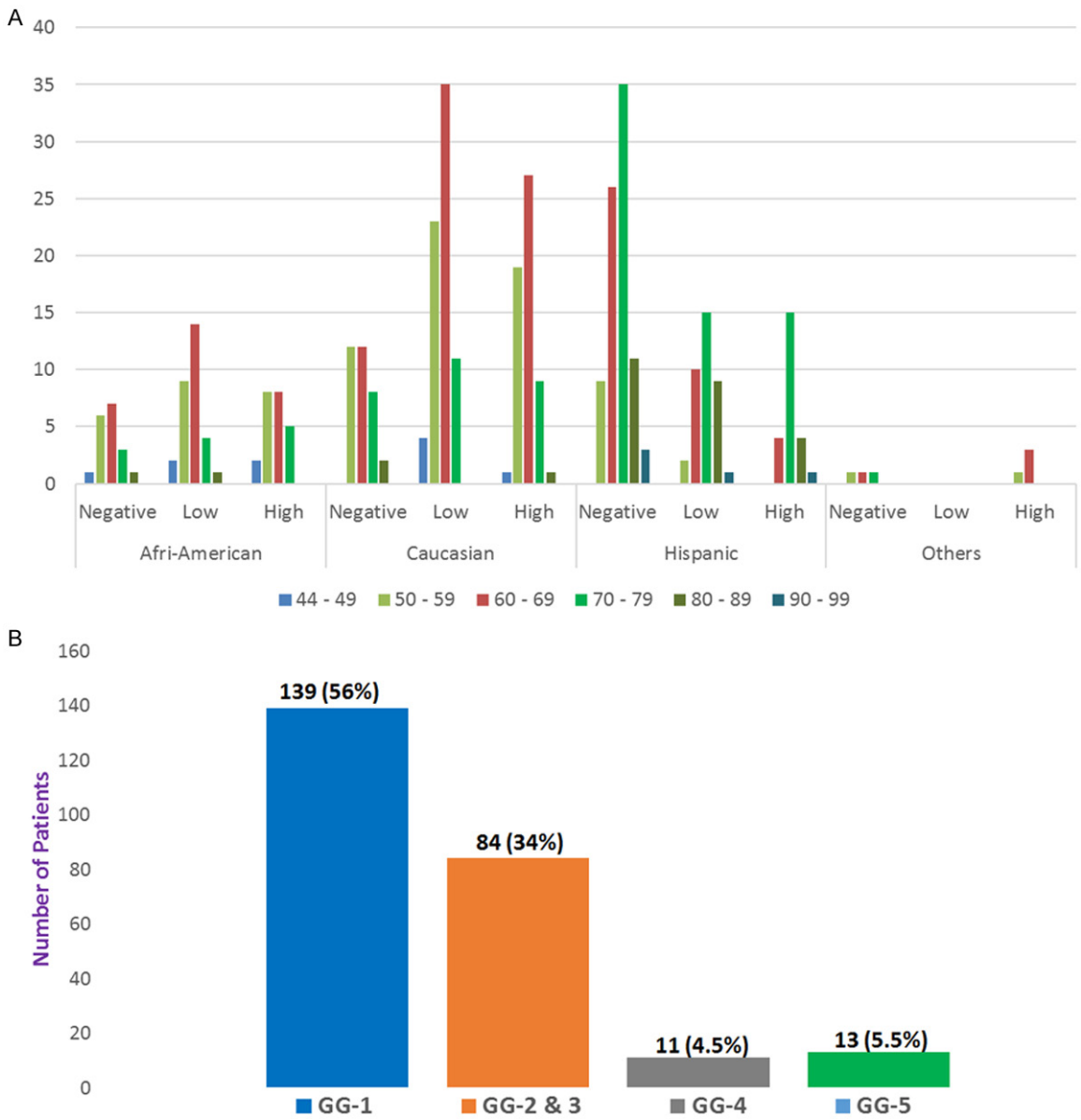

Figure 1. Sample demographic data. A: Age and racial distribution of control (negative), low-grade and intermediate/high-grade PCa urine samples. B: Gleason Grade Group (GG) distribution of biopsy-proven positive PCa patients for the risk assessment study.

showing the age and racial distribution of patients in these groups are summarized in **Table 1** and **Figure 1A**. The distribution of patients with biopsy-proven PCa in the Gleason group is shown in **Figure 1B**.

During the PCa modelling for the prediction of any PCa using the logistic regression machine learning model, 386 samples were randomly divided into two groups, with 67% and 33% of the data used for training and testing, respectively. Likewise, during PCa risk modelling to differentiate between low-grade and intermediate/high-grade PCa using the logistic regression model, the 247 PCa positive samples were randomly sub-divided into two groups, with 67% of the data used for training and 33% for testing.

Volatile organic compounds extraction from urine samples

To extract the VOCs from patients' urine samples, the -80°C stored urine samples were

thawed on ice at room temperature for approximately 30 min and centrifuged at 300 g for 10 min. Then, 1 mL of the supernatant was pipetted into a clean 20 mL amber vial, followed by the addition of 19 mL of DI H_2O, 600 µL of 2 M HCl, 300 µL of 1 ppm Mirex solution (internal standard), and a clean stir bar (TWISTER™, 10 mm × 1 mm, GERSTEL). The mixture was stirred at 1000 rpm for 2 h. Subsequently, the stir bar was removed from the solution, rinsed with DI H_2O, dried with lint-free tissue paper, and transferred into a thermal desorption tube (TDT). The tubes were then placed in an autosampler for VOC analysis. This VOC extraction method was described in our previous study assessing VOCs' performance of VOCs for predicting PCa in patients undergoing prostate biopsy [23].

Gas chromatography/mass spectrometry (GC-MS) coupled with thermal desorption

The VOC analysis was performed using an Agilent 8890 GC series system coupled with a mass spectrometer 5977B GC/MSD (Agilent Technologies, Delaware, USA). The system is in line with a Gerstel Multipurpose Sample (MPS) autosampler, Gerstel thermal desorption unit (TDU), and cooler injection system (CIS). Helium was used as the carrier gas at a constant flow rate of 1.0 mL/min. The Thermal desorption was programmed with an initial temperature of 45°C held for 30 s, then increased to 300°C and held for 5 min at 60°C/min. The resulting compounds released from the desorption process were collected into the CIS at -40°C, and the CIS was then heated to 300°C at 12°C/s and held for 5 min in splitless mode. The volatile compounds were then injected into the GC, before being separated and analyzed on an Agilent J&W HP - 5 ms Ultra Inert GC capillary column (30 m length × 0.25 mm internal diameter × 0.25 µm film thickness). The column oven temperature was held at 35°C for 5 min, then increased to 300°C, and held for 10 min at 10°C/min. The mass range was 20-500 *m/z*, and the data were gathered in the scan mode, as previously reported [23]. The Agilent Technologies GC-MS Enhanced Mass Hunter Workstation and Data Analysis Resource Application Software were used for programming this setup, as well as for data analysis. The NIST17 Library Search was used to identify the analyzed urinary volatile compounds present in the patient's urine samples. The NIST17 library search software identified each peak with the peak area and overall matching quality (%). A quality filter was then applied to remove any identified compounds with less than 50% matching quality.

Statistical data analysis

In modelling for the diagnosis of PCa and differentiation between low-grade and intermediate/high-grade PCa, where the outcomes are binary, an ultra-high dimensional classification challenge emerges. Similar approaches were employed for both types of analyses. To handle missing values, missing VOC values were imputed with zeroes. Variable screening was carried out using the two-sample t-test, simple logistic regression, and Wilcoxon rank-sum test. The results of the screening process were effectively visualized through volcano plots and heatmaps, providing valuable insights into the data. VOC variables that showed a significant difference (with *p*-value from Wilcoxon rank-sum test less than a threshold α) between the two statuses of outcome entered the subsequent modelling process.

To address the ultra-high dimensional classification problem, various machine-learning methods have been explored and compared using a repeated test-sample approach. Notably, regularized logistic regression consistently demonstrated superior performance in both PCa diagnosis and prognosis modelling. The final logistic model for the entire dataset was determined using the LASSO or l_1 penalty and fitted using Firth's approach. To assess the final model, a range of performance measures was reported based on jackknife-fitted values. These include the area under the receiver operating characteristic (ROC) curve (AUC), which reflects the sensitivity and specificity based on an optimal cut-off point. In addition, the important VOCs in the model were further explored. All the above-mentioned analyses were conducted using *R* and *MetaboAnalyst 5.0*, an *R*-based online open-source software developed in Canada.

Results

In this study, 386 urine samples were collected from different medical institutions across the United States for PCa screening and risk stratification in PCa biopsy pathologic assessment. The study samples were categorized based on

age brackets (44-99 years). In addition, racial diversity among the population was examined, and African Americans, Caucasians, Hispanics, and others were comparatively represented. The pathologic Gleason Grade Groups (GG) were also considered because this factor plays a critical clinical role in any PCa risk assessment and is one component of initial risk stratification. The Gleason Grade Groups of the samples ranged from GG-1 to GG-5, in addition to the PCa biopsy-proven negative samples, classified as negative controls. Details of the sample's demographics are provided in **Table 1** and **Figure 1A**, while **Figure 1B** indicates the Gleason group distributions among the respective low-grade (GG 1) and intermediate/high-grade (GG 2-5) PCa samples.

Extraction and identification of volatile organic compounds

To comprehensively determine and identify the corresponding VOCs that constitute the study, previously reported extraction and work-up protocols for urinary VOCs profiling procedures were employed [23]. The extracted VOCs were analyzed using GC-MS coupled with a thermal desorption unit in scan mode. Compounds were identified based on mass spectrometry measurements, and the abundance of each compound was determined with respect to instrument response signals based on the area under the peak. Mirex was used as an internal standard to calculate the relative abundance of each VOC extracted from the urine.

Using our published procedures [23], a total of 22,538 VOCs were identified in 386 urine samples. In the logistic regression Machine Learning (ML) prediction model for the diagnosis of any PCa, significant VOCs were determined and selected using the Wilcoxon Rank-sum test with a threshold of $P < 0.3$ and LASSO regression regularization for variable screening. The selected variables were used to generate ROC curves and the respective AUC values to determine the degree of sensitivity, specificity, and diagnostic accuracy. Furthermore, the raw data extracted from the NIST17 Library were separately subjected to PLS-DA multivariate analysis, a versatile algorithm used for predictive and descriptive modelling, and discriminative variable selection.

Logistic regression and partial least square discriminant analysis (PLS-DA)

The extracted urinary VOCs were analyzed for identification and quantification, and the data were subjected to a logistic regression machine learning algorithm and multivariate statistical analysis (PLS-DA). The logistic regression prediction model was explored. The diagnostic performance of differential VOC metabolites was evaluated using ROC curves, and the AUC values were computed as a measure of diagnostic accuracy to compare the groups. Likewise, PLS-DA was employed to explore the relationships between many different attributes because it can reduce data dimensionality, identify similarities, and discriminate samples with an extensive dataset without losing any vital information therein. Consequently, PLS-DA was employed in both the diagnosis model of PCa and the risk model for differentiation between low-risk and intermediate/high-risk PCa.

From the analysis, **Table 2A** depicts the VOCs selected by the Logistic regression model in PCa biopsy-proven positive and PCa biopsy-proven negative urine samples. **Table 2B** presents the top ten (10) most significant ($P < 0.05$) VOCs selected by the Wilcoxon rank-sum test in PCa biopsy positive and PCa biopsy negative urine samples with percentage occurrence and **Table 2C** represents the top ten (10) most significant ($P < 0.05$) VOCs selected by the Wilcoxon rank-sum test in low-grade and Intermediate/high-grade (referred to as high-grade in the table) PCa urine samples with percentage occurrence. The logistic regression model parameters such as the AUC values, percentages of sensitivity, specificity, and accuracy generated are presented in **Table 3**. From the data shown in **Table 3**, the logistic regression model predicted differences between patients negative for PCa and those with any PCa with 96% sensitivity, 76% specificity, and an accuracy level of 89% for the training set, and 92% sensitivity, 82% specificity, and an accuracy level of 87% for the testing set.

Discriminating between PCa biopsy-proven positive and negative urine samples: From the logistic regression analysis for the PCa diagnosis, ROC curves of training and testing sets of PCa biopsy-proven negative and positive PCa

Table 2. A: VOCs selected by Logistic regression model in PCa biopsy-proven positive and PCa biopsy-proven negative urine samples. B: The top ten (10) most significant (*P < 0.05*) VOCs selected by the Wilcoxon rank-sum test in PCa biopsy positive and PCa biopsy negative urine samples with percentage occurrence. C: The top ten (10) most significant (*P < 0.05*) VOCs selected by the Wilcoxon rank-sum test in low-grade and Intermediate/high-grade (herein referred to as high-grade in the table) PCa urine samples with percentage occurrence

A	Compound (CAS)	Chemical Name	Molecular Formula	*p*-values	PCa Status
1	000088-26-6	3,5-di-tert-Butyl-4-hydroxybenzyl alcohol	$C_{15}H_{24}O_2$	0.0002	PCa Positive
2	000063-42-3	Lactose	$C_{12}H_{22}O_{11}$	0.003	PCa Positive
3	000078-90-0	1,2-Propanediamine	$C_3H_{10}N_2$	0.01	PCa Positive
4	000205-25-4	7H-Benzo[c]carbazole	$C_{16}H_{11}N$	0.018	PCa Positive
5	000079-29-8	2,3-dimethylbutane	C_6H_{14}	0.019	PCa Positive
6	000115-77-5	2,2-bis(hydroxymethyl)propane-1,3-diol	$C_5H_{12}O_4$	0.038	PCa Positive
7	000107-21-1	Ethane-1,2-diol	$C_2H_6O_2$	0.038	PCa Positive
8	000083-47-6	Clionasterol	$C_{29}H_{50}O$	0.041	PCa Positive
9	000094-22-4	Propyl 4-nitrobenzoate	$C_{10}H_{11}NO_4$	0.007	PCa Negative
10	000054-91-1	Pipobroman	$C_{10}H_{16}Br_2N_2O_2$	0.013	PCa Negative

B	Compounds (CAS)	Chemical Name	Molecular Formula	*p*-value (P < 0.05)	% Occurrence	PCa Status
1	000098-86-2	Acetophenone	$C_6H_5COCH_3$	5.28×10^{-21}	79%	Positive
2	000540-97-6	Dodecamethylcyclohexasiloxane	$C_{12}H_{36}O_6Si_6$	3.55×10^{-19}	65%	Positive
3	001222-05-5	Galaxolide	$C_{18}H_{26}O$	1.04×10^{-18}	80%	Positive
4	000088-29-9	Versalide	$C_{18}H_{26}O$	1.24×10^{-18}	81%	Positive
5	000541-02-6	Decamethylcyclopentasiloxane	$C_{10}H_{30}O_5Si_5$	3.34×10^{-18}	65%	Positive
6	000629-50-5	Tridecane	$C_{13}H_{28}$	2.16×10^{-17}	79%	Positive
7	1000364-61-2	1,1,3,3,5,5,7,7-Octamethyl-7-(2-methylpropoxy)tetrasiloxan-1-ol	$C_{12}H_{34}O_5Si_4$	1.69×10^{-14}	69%	Positive
8	000101-86-0	Octanal, 2-(phenylmethylene)-	$C_{15}H_{20}O$	2.18×10^{-14}	94%	Positive
9	000112-31-2	Decanal	$C_{10}H_{20}O$	2.85×10^{-14}	64%	Positive
10	000872-05-9	1-Decene	$C_{10}H_{20}$	3.40×10^{-14}	67%	Positive

C	Compounds (CAS)	Chemical Name	Molecular Formula	*p*-value (P < 0.05)	% Occurrence	Risk Level
1	024569-83-3	Methyl Glyphosate	$C_4H_{10}NO_5P$	0.00059	61%	High-grade
2	024851-98-7	Methyl dihydrojasmonate	$C_{13}H_{22}O_3$	0.00098	68%	Low-grade
3	000057-56-7	Semicarbazide	CH_5N_3O	0.00174	66%	High-grade
4	000059-42-7	L-Phenylephrine	$C_9H_{13}NO_2$	0.00181	73%	Low-grade
5	042775-75-7	5-ethyl-1,2,3,4-tetrahydronaphthalene	$C_{12}H_{16}$	0.00241	100%	High-grade
6	1000417-21-0	Norcodeine, N-trimethylsilyl-, trimethylsilyl ether	$C_{23}H_{35}NO_3Si_2$	0.0024	100%	High-grade
7	002460-77-7	2,5-Di-tert-butyl-1,4-benzoquinone	$[(CH_3)_3C]_2C_6H_2(=O)_2$	0.00371	57%	High-grade
8	114274-97-4	2-phenyl-3-decyn-1-ol	$C_{16}H_{22}O$	0.00456	100%	Low-grade
9	000110-63-4	Butane-1,4-diol	$HO(CH_2)_4OH$	0.00505	100%	High-grade
10	007087-68-5	N,N-Diisopropylethylamine	$C_8H_{19}N$	0.00507	74%	Low-grade

were generated as shown in **Figure 2A**. The AUC diagnosis and prediction values for the training and testing datasets were 0.99 and 0.88, respectively. This shows that the model has a high prediction accuracy in differentiating between patients without PCa and patients with PCa.

In addition, the PLS-DA analysis score plot was generated to discriminate between PCa biopsy-proven negative and positive patients **Figure 2B**. To further explain the variance observed in the PLS-DA score plot, the variable importance in projection (VIP) loading plot of the top 25 most significant VOCs with their corresponding CAS number was provided in **Figure 2B**. The Variable Importance in Projection (VIP) scores estimate the importance of each variable in the projection used in a PLS-DA model and are often used for variable selection. VIP is a weighted sum of squares of the PLS-DA loadings considering the amount of explained Y-variable in each dimension. A variable with a VIP Score close to or greater than 1 can be con-

Table 3. AUC values, percentages of sensitivity, specificity and accuracy from Logistic regression models

Groups	Sample-Set	AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)
PCa Positive vs Negative	Training	0.99	96%	76%	89%
	Testing	0.88	92%	82%	87%
Low-grade vs Intermediate/High-grade	Cross-validation	0.78	80%	70%	75%

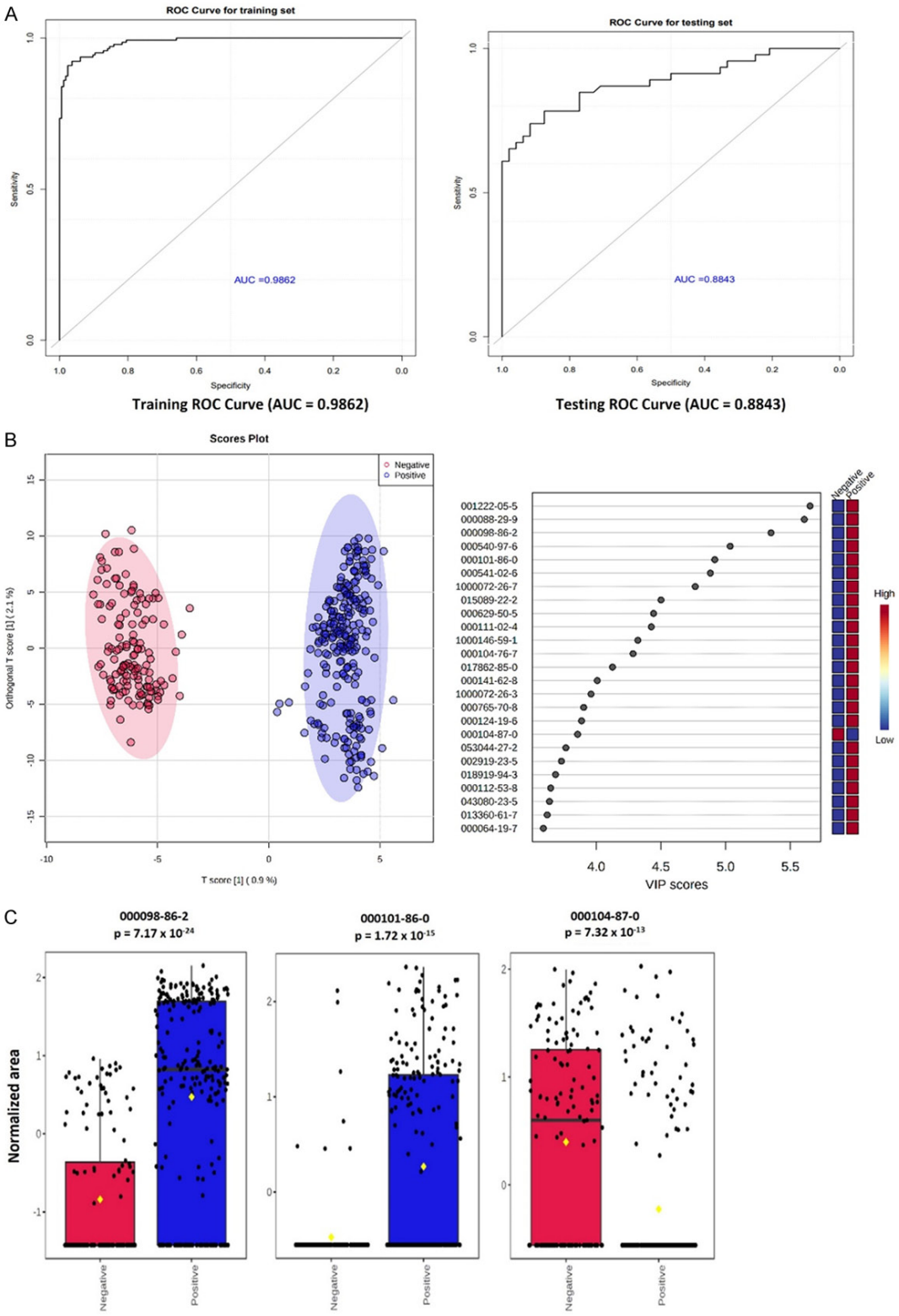

Figure 2. A: ROC curves of training and testing sets of PCa biopsy-proven negative and positive PCa. PLS-DA plots of PCa biopsy-proven negative and positive patients. B: The variable importance in projection (VIP) loading plot of the top 25 most significant VOCs (indicated in their corresponding CAS numbers). The colored boxes on the right of the PLS-DA score plot indicate the relative concentrations of the corresponding metabolites in each group under study. C: Boxplots of some of the significant VOCs in discriminating PCa biopsy-proven negative and positive urine samples.

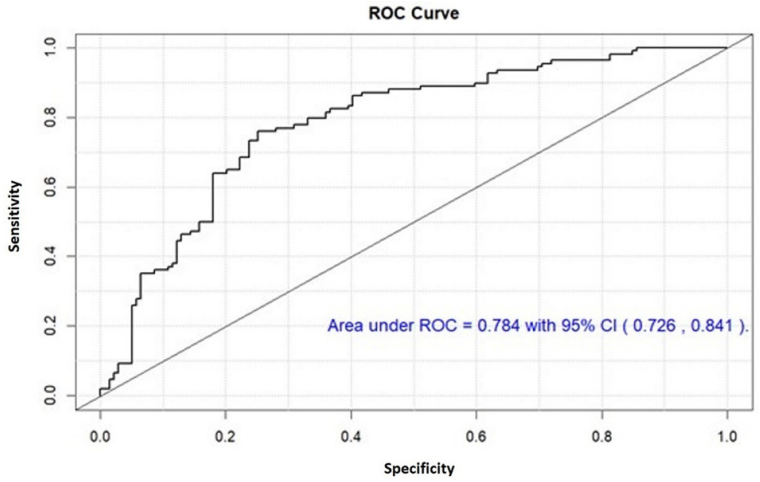

Figure 3. ROC curve of training and testing sets of low-grade and intermediate/high-grade PCa based on jackknife prediction of LASSO regularized logistic regression model with tuning parameter of 0.027 (mean) and AUC value of 0.78.

sidered important in the model. The VIP important features (i.e., VOCs) identified by PLS-DA in descending order of importance. The plot represents the relative contribution of the VOCs to the variance between the PCa biopsy-proven negative and positive urine samples. The colored boxes on the right of the PLS-DA score plot indicate the relative concentrations of the corresponding metabolites in each group under study. Furthermore, boxplots of some of the top significant VOCs in discriminating PCa biopsy-proven negative and positive urine samples were provided in **Figure 2C** to show the differences in the relative concentrations of some significant VOCs when compared between the two groups.

Discriminating between low-grade PCa and intermediate/high-grade PCa: Moreover, the same methodology used for the PCa diagnosis was explored in the risk assessment (prognosis) study. Likewise, the ROC curve was generated for the training and testing sets of low-grade and intermediate/high-grade PCa based on jackknife prediction of LASSO regularized logistic regression model with tuning parameter of 0.027 (mean) and AUC value of 0.78 as shown in **Figure 3**.

Likewise, the PLS-DA multivariate statistical algorithm was explored (as applied in the PCa diagnosis section) using the raw dataset extracted from the NIST17 Library. **Figure 4A** illustrates the PLS-DA score plot of the low-grade and intermediate/high-grade urine samples, whereas the VIP loading plot in **Figure 4B** shows the representative biomarkers (25 out of 123 significant VOCs) (information will be available on request from the corresponding authors) responsible for the distinction in the score plot. The VOCs were selected based on their *p*-values significant ($P < 0.05$) by the Wilcoxon rank-sum test analysis with the percentage (%) of occurrence to discriminate between the low-grade and intermediate/high-grade urine samples (**Table 2C**). These results suggest that PLS-DA can also be employed for the prediction of intermediate/high-grade cancer among patients with PCa. This could be helpful in monitoring the disease of patients diagnosed with low-grade PCa on active surveillance to help determine the timing of repeat prostate biopsies. In addition, **Figure 4C** shows some selected significant VOCs that contributed to the separation observed in the PLS-DA score and VIP loading plots.

Discussion

Serum, tissue, and urinary sample matrices have been investigated in metabolomics as potential cancer screening techniques [24-26]. Urine is the most common biospecimen collected during routine testing as it comprises a significant source of organic metabolic compounds [27]. Studies reported that dogs could detect cancers by smelling the urine from patients with prostate [28], bladder [29], breast and lung cancer [30]. Those studies highlight-

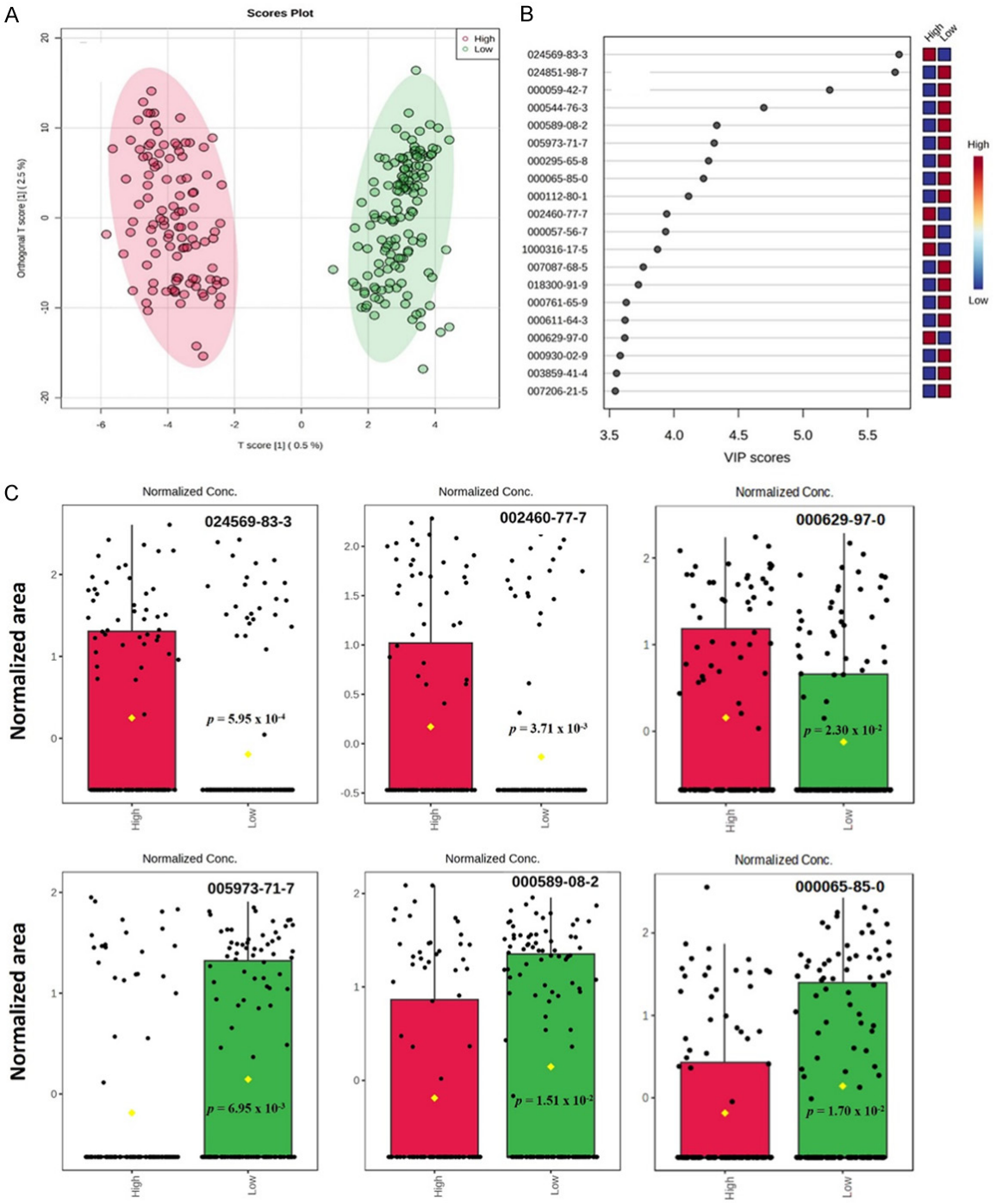

Figure 4. A: PLS-DA score plot of biopsy-proven positive low-grade and intermediate/high-grade PCa. B: The VIP loading plot represents the variable importance in projection (VIP) of each metabolite, while the colored boxes on the right of the PLS-DA score plot indicate the relative concentrations of the corresponding metabolite in each group under study. C: Examples of boxplots of significant VOCs in PCa low-grade and intermediate/high-grade urine samples.

ed the applications of urinary VOCs for cancer detection.

VOCs are a diverse carbon-based group of chemical compounds that reflect metabolomics and are suspected to be predictive of PCa status. Importantly, VOCs can serve as pathogen identifiers and invariably as disease signature biomarkers. Analysis of urinary VOCs may be able to provide a fast and non-invasive diag-

nostic method to help identify patients with PCa as well as differentiate between patients with low-grade PCa and intermediate/high-grade PCa.

Discriminate between PCa biopsy-proven positive and negative urine samples

Logistic regression modelling analysis for PCa diagnosis: In the first part of this study, we investigated the difference in urinary VOCs between PCa-positive and -negative patients using samples from 139 PCa biopsy-negative and 247 PCa biopsy-positive patients. To avoid omitting important predictors for the presence of PCa, a relatively high threshold ($P < 0.3$) was applied to screen variables to develop the regression model, resulting in a total of 2,354 potential VOCs. After *L1* regularization, the final logistic model for predicting the presence of PCa comprised 10 VOCs, as listed in **Table 2A**. These 10 VOCs are considered significant based on the model analysis outcome and also possess the potential to distinguish between PCa-positive and PCa-negative patients.

To assess the performance of this predictive model, ROC curves were generated for the training and testing datasets are shown in **Figure 2A**. The ROC (receiver operating characteristic) curve is a plot illustrating the performance of a classification model. It is a binary classifier model at varying threshold values. The curve plots the true positive rate against the false positive rate at each threshold setting. In other words, it is a probability curve, and the AUC represents the degree or measure of separability. It explains how much the model is capable of distinguishing between different groups. In using this model for diagnosis, an AUC of 0.5 suggests no discrimination (i.e., no ability to diagnose patients with and without the disease or condition based on the test), 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, more than 0.9 is considered outstanding and 1 indicates perfect performance. AUC is mostly considered because it is the most reliable metric that calibrates the trade-off between sensitivity and specificity at the best-chosen threshold, or cut-off point. Accuracy measures how well a single model is doing, whereas AUC compares two models as well as evaluates the same model's performance across different thresholds. **Figure 2A** illustrates that the AUC diagnosis and prediction values for the training and testing datasets were 0.99 and 0.88, respectively. This shows that this diagnostic model has a high prediction accuracy in differentiating between patients without PCa and patients with PCa.

Partial least square-discriminant analysis for PCa diagnosis: Additionally, the PLS-DA analytical method was used to discriminate between samples from patients negative for PCa and patients with any PCa. From the PLS-DA multivariate analysis, PCa positive and PCa negative were easily distinguished in the score plots, as illustrated in **Figure 2B**. The variable importance in projection (VIP) score measures the importance of a variable in the PLS-DA statistical model. The score sums up the contributions of a variable to the model. The X-axis indicates the VIP scores corresponding to each variable (i.e., VOC) on the y-axis, and the black dots in the loading plot indicate the factors with the highest VIP scores and thus are the most contributory variables in class discrimination in the PLS-DA model. There were 229 VOCs with VIP scores greater than 1 (information will be available on request from the corresponding authors), although only 25 are shown in **Figure 2B** score plot to demonstrate the findings. The VIP loading plot in **Figure 2B** shows that most of the significant 25 VOCs for the predictive model were elevated in the PCa-positive group but were downregulated in the PCa-negative group. **Figure 2C** also depicts some examples of the different VOC concentration levels between PCa positive and negative groups, while **Table 2B** depicts the top ten (10) most significant VOCs that were selected based on their *p*-values ($P < 0.05$) by the Wilcoxon rank-sum test analysis in positive PCa biopsy-positive and negative urine samples with the percentage (%) of occurrence (**Table 2B**). These VOCs could be potential biomarkers for non-invasive differentiation between patients without PCa and patients with PCa. For patients who are unlikely to have PCa based on biomarkers, this could help avoid unnecessary invasive prostate biopsies and the associated harms. Conversely, for patients found to be highly likely to have PCa based on biomarkers, this could help patients better understand their risk of having PCa and make a better decision on whether to undergo prostate biopsy.

Discriminate between low-grade PCa and intermediate/high-grade PCa

This study further explored the use of VOCs in PCa to differentiate between patients with low- and intermediate/high-grade PCa. As an exploratory study, we categorized the 247 PCa biopsy-proven positive patients' urine samples into two sub-groups: 139 were categorized as low-grade with an ISUP Gleason Grade Group 1, and 108 were classified as intermediate/high-grade with an ISUP Gleason Grade Group 2 and above (**Table 1** and **Figure 1B**).

Logistic regression modelling analysis for PCa prognosis: For variable screening, we noticed that the predictive power of individual VOCs in modelling differentiation between low-grade PCa and intermediate/high-grade PCa is significantly weaker (information will be available on request from the corresponding authors). To identify the most effective approach, we partitioned the data into training and test samples with a ratio of 70:30 and conducted experiments with a variety of machine learning (ML) algorithms, including regularized logistic regression with different types of penalties, partial least squares, decision trees, support vector machines (SVM), kernel KNN, random forests (RF), naïve Bayes classifier, and boosting. In addition, two thresholds of α (α = 0.20 and α = 0.05) were assessed for variable screening. The performance of each approach was evaluated using the AUC measure based on the test sample. According to the results, regularized logistic regression with LASSO or elastic net demonstrated strong performance (information will be available on request from the corresponding authors). An intriguing observation was that using a smaller α threshold resulted in higher AUC values across all methods (information will be available on request from the corresponding authors). This can be attributed to the fact that a smaller α filters out more variables, thus simplifying the prediction task. Considering that LASSO enables variable selection, we decided to focus our efforts on further exploring regularized logistic regression with the LASSO penalty as the chosen approach.

To enhance the reliability of our evaluation, we conducted experiments using regularized logistic regression with LASSO penalty, employing 100 runs of random data partitioning. The average AUC obtained from these experiments, with a threshold of α = 0.05, was calculated to be 0.78 (Standard Deviation (SD) of 0.095), indicating good predictive performance. Conversely, at α = 0.20, the average AUC was 0.52 (SD 0.059), highlighting a significant drop in performance. With over 100 runs, we obtained the optimal selected tuning parameter λ values from a histogram plot with a mean of 0.027 and an SD of 0.024 (detailed information will be available on request from the corresponding authors).

Consequently, our final model for the entire dataset was a regularized logistic regression model that includes 97 VOCs (information will be available on request from the corresponding authors) employing an α threshold of 0.05 for variable screening and a tuning parameter of λ = 0.027 for the LASSO penalty. To obtain the standard error (SE) and handle the complete separation, the final model fitting was performed using Firth's approach. To evaluate the performance of the final model, a Jackknife approach was employed to compute the leave-one-out predicted probability of each subject in the intermediate/high-grade PCa category. The resulting area under the curve (AUC) was 0.78 (with 95% confidence intervals of 0.726 and 0.841) (**Figure 3**). It demonstrated that this model prognosis prediction rate with the AUC value of 0.78 is good in discriminating between low-grade and intermediate/high-grade PCa. Furthermore, as shown in **Table 3**, the logistic regression model had sensitivity (80%) and specificity (70%) in predicting the difference between patients with low-grade and intermediate/high-grade PCa.

However, it was observed that the predictive power of our model differentiating between low-grade PCa and intermediate/high-grade PCa prognosis is significantly weaker (**Figure 3**) than what was demonstrated in the prediction of PCa diagnosis (**Figure 2A**). It could be due to the fact that PCa is a heterogeneous disease and that the variations among indolent low-grade tumors to large aggressive life-threatening tumors could not be simplified as binary outcomes. For example, GG2 cancer is Gleason 3 + 4. For that diagnosis, the amount of pattern 4 can vary from minimal, e.g., 5%, to approaching 50%. The greater the presence of pattern 4 in the GG 2 diagnosis, the more indicative it

becomes of unfavorable clinical significance and risk; while GG 2 with less than 20% pattern 4 behave like GG 1. Thus, it could be difficult for a binary algorithm diagnostic model to distinctly separate or discriminate between low-grade and intermediary/high-grade samples during PCa prognosis compared to when the model is used for diagnosis.

Partial least square-discriminant analysis for PCa prognosis: Among the 123 significant VOCs found in the PLS-DA multivariate statistical analysis for the PCa risk stratification model, 74 VOCs were elevated in intermediate/high-grade PCa compared with low-grade PCa, while 49 VOCs were decreased in intermediate/high-grade PCa relative to patients with low-grade PCa (information will be available on request from the corresponding authors). The top ten significant VOCs selectede by PLS-DA, methyl glyphosate ($C_4H_{10}NO_5P$), semicarbazide (CH_5N_3O), and 5-ethyl-1,2,3,4-tetrahydronaphthalene ($C_{12}H_{16}$), with *p*-values 5.95×10^{-4}, 1.70×10^{-3} and 2.40×10^{-3} respectively, are significantly elevated in intermediate/high-grade PCa group; while methyl dihydrojasmonate ($C_{13}H_{22}O_3$) and L-phenylephrine ($C_9H_{13}NO_2$) are significantly elevated for the low-grade PCa group, with *p*-values 9.76×10^{-4} and 1.80×10^{-3} respectively. These VOCs could be explored as clinical biomarkers for the identification of patients with intermediate/high-grade PCa to help decrease unnecessary prostate biopsies and decrease the risk of over-diagnosis and over-treatment of low-grade PCa patients. Additionally, these VOCs may be further validated and explored as clinical biomarkers for monitoring patients diagnosed with low-grade PCa during active surveillance to help determine the optimal timing for further intervention, including repeating prostate biopsies or changing to active treatment.

It should be noted that this study implemented non-targeted metabolomics to screen for metabolites that are significantly different between the urine of PCa patients and cancer-negative controls, as well as between the urine of low grade PCa and of intermediate/high grade PCa. Significant VOCs were identified using statistical analysis (i.e., Wilcoxon Rank-Sum Test) and machine learning (i.e., Logistic Regression and PLS-DA). Thus, the selection of those VOCs was unbiased and without any specific pathway considerations.

As shown in **Table 2**, many significant VOCs, such as 3,5-di-tert-Butyl-4-hydroxybenzyl alcohol, 2,3-dimethylbutane, Acetophenone, Tridecane, Decanal, and 1-Decene are hydrocarbons, ketones and aldehydes that could be due to oxidative stress, which is a common denominator in the pathogenesis of cancer and other chronic diseases [31].

Many compounds selected by the logistic regression model and PLS-DA were related to products of agricultural or industrial applications. Although it is out of the scope of this study to examine the potential source of these VOCs detected in urine, we further examined the biological significance of several significant VOCs specifically found in differentiating low- and intermediate/high-grade prostate cancer (**Table 2C**).

Significant volatile organic compounds of biological importance

Methyl Glyphosate was found to be the most significant VOC predominantly in high-grade PCa (P = 0.00059). It was also an unexpected urinary metabolite found in the samples as it is an herbicide. It is known that Glyphosate is a broad-spectrum, non-selective, post-emergent, systemic herbicide that kills or suppresses all plant types, apart from those genetically modified and designed to resist glyphosate [32]. Glyphosate is among the most utilized herbicides in the United States and the second most utilized weedkiller in agriculture, gardens, and industry. Although the U.S. Environmental Protection Agency (EPA) reported no evidence that glyphosate causes cancer in humans, the cytotoxicity and oxidative stress responses of glyphosate in a human prostate study demonstrated that glyphosate is a very toxic in vitro assay with the potential ability to induce apoptotic effects as well as oxidative stress [33]. Glyphosate is an organophosphate compounds and exposure to organophosphate insecticides has been linked to prostate cancer in many studies, where it has been reported that organophosphate insecticides are significantly associated with aggressive prostate cancer [34]. Recently, concerns have been raised about human exposure to glyphosate following an assessment by the International Agency for Research on Cancer (IARC) that re-classified glyphosate as 'probably carcinogenic to humans' [35]. Additionally, occupational exposure

to pesticides increases the risk of prostate and other cancer types [36-38].

Semicarbazide was found to be the 3rd significant compound in the PLS-DA analysis. It is a minor thermal decomposition product of azodicarbonamide employed to produce gaskets of food jars [39]. Semicarbazide inhibits semicarbazide-sensitive amine oxidase. Research found that the level of specific semicarbazide-sensitive amine oxidase activity in serum was significantly elevated in patients with skeletal metastases as compared to both controls and PCa patients without signs of skeletal metastases [40]. Semicarbazide was found to modify testicular morphology in male zebrafish and lower the somatic testicular index, ultimately reducing reproductive regulation. Likewise, semicarbazide has been reported to inhibit the mediated effect of glutamic acid decarboxylase (GAD) and gamma-aminobutyric acid (GABA), which have been tested and validated in the reproductive system of male zebrafish, to cause disorders via GABA in males [41, 42]. Overproduction of GABA can directly regulate nuclear androgen receptor (AR) signalling to drive prostate cancer [43, 44]. Taking into consideration that semicarbazide interferences with GABA system [45], the role of semicarbazide as biomarkers could be further investigated.

The clinical implication of 5-ethyl-1,2,3,4-tetrahydronaphthalene in association with prostate cancer is unknown. Generally, naphthalene is a polycyclic aromatic hydrocarbon associated with health effects, including cancer. Naphthalene and its metabolites are harmful to humans. Naphthalene induces respiratory tract toxicity in mice [46].

Methyl dihydrojasmonate and l-phenylephrine levels were significantly associated with low-grade PCa. Based on several studies conducted on methyl dihydrojasmonate, there is currently no data on the carcinogenic, mutagenic, or genotoxic potential of this compound [47, 48]. Phenylephrine is mainly used as a vasoconstrictor and is not usually recommended for regular prescriptions. However, it is used in acute medical conditions to increase mean arterial pressure in the presence of high cardiac output and acute tachyarrhythmias induced by norepinephrine [49]. Thus, it could be suggested that these two VOCs may not be directly related to risk in PCa but probably to PCa patients' anticancer drug metabolites due to prostate cancer management.

Potential clinical implications

Considering the low specificity of the current PCa diagnosis method (serum PSA screening), more than 80% of men with elevated PSA concentration often have to undergo PCa biopsy; and they will either not be diagnosed with PCa or be diagnosed with an indolent form of PCa, which invariably might not lead to morbidity or mortality. Thus, it is of essence to develop a more accurate, precise, and fast method for PCa diagnosis and prognosis to prevent the challenges associated with overdiagnosis and overtreatment of PCa. With this study, a more accurate method has been developed for PCa diagnosis (AUC = 0.88) and prognosis (AUC = 0.78) with the use of VOCs metabolomics methodology which has the potential of providing a good alternative to the current method for PCa diagnosis and prognosis. Differentiating between low-grade PCa and intermediate/high-grade PCa could also be applied to disease monitoring of patients with low-risk PCa on active surveillance for further intervention when needed. This reported methods are non-invasive, faster, and more accurate, and it could also be a more cost effective option for the PCa patients for both diagnosis and prognosis.

Conclusion

In this study, we established a panel of urinary VOC biomarkers that could potentially help distinguish between patients without PCa and those with PCa, as well as differentiate between patients with low-grade PCa and intermediate/high-grade PCa. Urinary VOC biomarkers are non-invasive and cost-effective for use in the diagnosis of PCa and disease monitoring in patients with low-grade PCa. When used as a tool for helping with PCa diagnosis, the use of these biomarkers could help decrease the need for diagnostic prostate biopsies, decrease the problems of over-diagnosis and overtreatment of low-grade PCa, and improve patients' understanding of their risk of PCa prior to deciding to pursue a prostate biopsy. In addition, when used as a tool for monitoring disease progression in patients with low-grade PCa, the use of these biomarkers could help

determine the optimal timing for repeat prostate biopsies for patients undergoing active surveillance to prevent unnecessary repeat prostate biopsies and more quickly identify patients with PCa that have become more aggressive. Hence, upon further clinical validation, this urinary VOC assessment method could provide improved outcomes in the identification and management of patients with PCa.

Acknowledgements

The authors acknowledge the contribution of the East Virginia Medical School Biorepository in providing some of the patients' urine samples used in this study. The research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health (NCI-NIH) under award numbers SC1CA245675, U54MD007592, and 1T32GM144919. The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure of conflict of interest

None.

Address correspondence to: Wen-Yee Lee, Department of Chemistry and Biochemistry, University of Texas at El Paso, 500 W University Road, El Paso, Texas 79968, USA. E-mail: wylee@utep.edu

References

[1] Descotes JL. Diagnosis of prostate cancer. Asian J Urol 2019; 6: 129-136.
[2] Tonry C, Finn S, Armstrong J and Pennington SR. Clinical proteomics for prostate cancer: understanding prostate cancer pathology and protein biomarkers for improved disease management. Clin Proteomics 2020; 17: 41.
[3] Siegel RL, Miller KD, Fuchs HE and Jemal A. Cancer statistics, 2022. CA Cancer J Clin 2022; 72: 7-33.
[4] Chan JM, Gann PH and Giovannucci EL. Role of diet in prostate cancer development and progression. J Clin Oncol 2005; 23: 8152-8160.
[5] Rawla P. Epidemiology of prostate cancer. World J Oncol 2019; 10: 63-89.
[6] Sharma S, Cwiklinski K, Sykes DE, Mahajan SD, Chevli K, Schwartz SA and Aalinkeel R. Use of glycoproteins-prostate-specific membrane antigen and galectin-3 as primary tumor markers and therapeutic targets in the management of metastatic prostate cancer. Cancers (Basel) 2022; 14: 2704.
[7] Desai MM, Cacciamani GE, Gill K, Zhang J, Liu L, Abreu A and Gill IS. Trends in incidence of metastatic prostate cancer in the US. JAMA Netw Open 2022; 5: e222246.
[8] Humphrey PA. Histopathology of prostate cancer. Cold Spring Harb Perspect Med 2017; 7: a030411.
[9] Qiu S, Cai Y, Yao H, Lin C, Xie Y, Tang S and Zhang A. Small molecule metabolites: discovery of biomarkers and therapeutic targets. Signal Transduct Target Ther 2023; 8: 132.
[10] Castelli FA, Rosati G, Moguet C, Fuentes C, Marrugo-Ramírez J, Lefebvre T, Volland H, Merkoçi A, Simon S, Fenaille F and Junot C. Metabolomics for personalized medicine: the input of analytical chemistry from biomarker discovery to point-of-care tests. Anal Bioanal Chem 2022; 414: 759-789.
[11] Schmidt DR, Patel R, Kirsch DG, Lewis CA, Vander Heiden MG and Locasale JW. Metabolomics in cancer research and emerging applications in clinical oncology. CA Cancer J Clin 2021; 71: 333-358.
[12] Clish CB. Metabolomics: an emerging but powerful tool for precision medicine. Cold Spring Harb Mol Case Stud 2015; 1: a000588.
[13] Gomez-Cebrian N, Rojas-Benedicto A, Albors-Vaquer A, Lopez-Guerrero JA, Pineda-Lucena A and Puchades-Carrasco L. Metabolomics contributions to the discovery of prostate cancer biomarkers. Metabolites 2019; 9: 48.
[14] Gowda GA, Zhang S, Gu H, Asiago V, Shanaiah N and Raftery D. Metabolomics-based methods for early disease diagnostics. Expert Rev Mol Diagn 2008; 8: 617-633.
[15] Shockcor JP and Holmes E. Metabonomic applications in toxicity screening and disease diagnosis. Curr Top Med Chem 2002; 2: 35-51.
[16] Kdadra M, Hockner S, Leung H, Kremer W and Schiffer E. Metabolomics biomarkers of prostate cancer: a systematic review. Diagnostics (Basel) 2019; 9: 21.
[17] Kuehnbaum NL and Britz-McKibbin P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. Chem Rev 2013; 113: 2437-2468.
[18] Wishart DS. Metabolomics for investigating physiological and pathophysiological processes. Physiol Rev 2019; 99: 1819-1875.
[19] Cornu JN, Cancel-Tassin G, Ondet V, Girardet C and Cussenot O. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. Eur Urol 2011; 59: 197-201.
[20] Giro Benet J, Seo M, Khine M, Guma Padro J, Pardo Martnez A and Kurdahi F. Breast cancer detection by analyzing the volatile organic compound (VOC) signature in human urine. Sci Rep 2022; 12: 14873.

[21] Khalid T, Aggio R, White P, De Lacy Costello B, Persad R, Al-Kateb H, Jones P, Probert CS and Ratcliffe N. Urinary volatile organic compounds for the detection of prostate cancer. PLoS One 2015; 10: e0143283.
[22] Nakhleh MK, Amal H, Jeries R, Broza YY, Aboud M, Gharra A, Ivgi H, Khatib S, Badarneh S, Har-Shai L, Glass-Marmor L, Lejbkowicz I, Miller A, Badarny S, Winer R, Finberg J, Cohen-Kaminsky S, Perros F, Montani D, Girerd B, Garcia G, Simonneau G, Nakhoul F, Baram S, Salim R, Hakim M, Gruber M, Ronen O, Marshak T, Doweck I, Nativ O, Bahouth Z, Shi DY, Zhang W, Hua QL, Pan YY, Tao L, Liu H, Karban A, Koifman E, Rainis T, Skapars R, Sivins A, Ancans G, Liepniece-Karele I, Kikuste I, Lasina I, Tolmanis I, Johnson D, Millstone SZ, Fulton J, Wells JW, Wilf LH, Humbert M, Leja M, Peled N and Haick H. Diagnosis and classification of 17 diseases from 1404 subjects via pattern analysis of exhaled molecules. ACS Nano 2017; 11: 112-125.
[23] Gao Q, Su X, Annabi MH, Schreiter BR, Prince T, Ackerman A, Morgas S, Mata V, Williams H and Lee WY. Application of urinary volatile organic compounds (VOCs) for the diagnosis of prostate cancer. Clin Genitourin Cancer 2019; 17: 183-190.
[24] Pastore AL, Palleschi G, Silvestri L, Moschese D, Ricci S, Petrozza V, Carbone A and Di Carlo A. Serum and urine biomarkers for human renal cell carcinoma. Dis Markers 2015; 2015: 251403.
[25] Nishiumi S, Shinohara M, Ikeda A, Yoshie T, Hatano N, Kakuyama S, Mizuno S, Sanuki T, Kutsumi H and Fukusaki E. Serum metabolomics as a novel diagnostic approach for pancreatic cancer. Metabolomics 2010; 6: 518-528.
[26] Naz S, Moreira dos Santos DC, García A and Barbas C. Analytical protocols based on LC-MS, GC-MS and CE-MS for nontargeted metabolomics of biological tissues. Bioanalysis 2014; 6: 1657-1677.
[27] Abate-Shen C and Shen MM. Diagnostics: the prostate-cancer metabolome. Nature 2009; 457: 799-800.
[28] Cornu JN, Cancel-Tassin G, Ondet V, Girardet C and Cussenot O. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. Eur Urol 2011; 59: 197-201.
[29] Willis CM, Britton LE, Harris R, Wallace J and Guest CM. Volatile organic compounds as biomarkers of bladder cancer: sensitivity and specificity using trained sniffer dogs. Cancer Biomark 2010; 8: 145-153.
[30] McCulloch M, Jezierski T, Broffman M, Hubbard A, Turner K and Janecki T. Diagnostic accuracy of canine scent detection in early- and late-stage lung and breast cancers. Integr Cancer Ther 2006; 5: 30-39.
[31] Pirrone F and Albertini M. Olfactory detection of cancer by trained sniffer dogs: a systematic review of the literature. J Vet Behav 2017; 19: 105-117.
[32] Tarazona JV, Court-Marques D, Tiramani M, Reich H, Pfeil R, Istace F and Crivellente F. Glyphosate toxicity and carcinogenicity: a review of the scientific basis of the European Union assessment and its differences with IARC. Arch Toxicol 2017; 91: 2723-2743.
[33] Abdel-Halim KY and Osman SR. Cytotoxicity and oxidative stress responses of imidacloprid and glyphosate in human prostate epithelial WPM-Y.1 cell line. J Toxicol 2020; 2020: 4364650.
[34] Pardo LA, Beane Freeman LE, Lerro CC, Andreotti G, Hofmann JN, Parks CG, Sandler DP, Lubin JH, Blair A and Koutros S. Pesticide exposure and risk of aggressive prostate cancer among private pesticide applicators. Environ Health 2020; 19: 30.
[35] Davoren MJ and Schiestl RH. Glyphosate-based herbicides and cancer risk: a post-IARC decision review of potential mechanisms, policy and avenues of research. Carcinogenesis 2018; 39: 1207-1215.
[36] Alavanja MC and Bonner MR. Occupational pesticide exposures and cancer risk: a review. J Toxicol Environ Health B Crit Rev 2012; 15: 238-263.
[37] Koutros S, Berndt SI, Hughes Barry K, Andreotti G, Hoppin JA, Sandler DP, Yeager M, Burdett LA, Yuenger J, Alavanja MC and Beane Freeman LE. Genetic susceptibility loci, pesticide exposure and prostate cancer risk. PLoS One 2013; 8: e58195.
[38] Weisenburger DD. A review and update with perspective of evidence that the herbicide glyphosate (Roundup) is a cause of non-hodgkin lymphoma. Clin Lymphoma Myeloma Leuk 2021; 21: 621-630.
[39] Stadler RH, Mottier P, Guy P, Gremaud E, Varga N, Lalljie S, Whitaker R, Kintscher J, Dudler V, Read WA and Castle L. Semicarbazide is a minor thermal decomposition product of azodicarbonamide used in the gaskets of certain food jars. Analyst 2004; 129: 276-281.
[40] Ekblom J, Grönvall J, Lennernäs B, Nilsson S, Garpenstrand H and Oreland L. Elevated activity of semicarbazide-sensitive amine oxidase in blood from patients with skeletal metastases of prostate cancer. Clin Sci (Lond) 1999; 97: 111-115.
[41] Tian X, Li H, Zhang X, Xu Y, Zhang H, Han D, Liu H, Wang B, Cui Y, Liu H, Zhou Q and Gong X. Effects of acute and chronic exposure to semi-

carbazide on the sea cucumber apostichopus japonicus. Front Environ Sci 2021; 9.
[42] Yu M, Feng Y, Zhang X, Wang J, Tian H, Wang W and Ru S. Semicarbazide disturbs the reproductive system of male zebrafish (Danio rerio) through the GABAergic system. Reprod Toxicol 2017; 73: 149-157.
[43] Taylor RA and Watt MJ. Unsuspected protumorigenic signaling role for the oncometabolite GABA in advanced prostate cancer. Cancer Res 2019; 79: 4580-4581.
[44] Solorzano SR, Imaz-Rosshandler I, Camacho-Arroyo I, García-Tobilla P, Morales-Montor G, Salazar P, Arena-Ortiz ML and Rodríguez-Dorantes M. GABA promotes gastrin-releasing peptide secretion in NE/NE-like cells: contribution to prostate cancer progression. Sci Rep 2018; 8: 10272.
[45] Maranghi F, Tassinari R, Marcoccia D, Altieri I, Catone T, De Angelis G, Testai E, Mastrangelo S, Evandri MG, Bolle P and Lorenzetti S. The food contaminant semicarbazide acts as an endocrine disrupter: evidence from an integrated in vivo/in vitro approach. Chem Biol Interact 2010; 183: 40-48.
[46] Buckpitt A, Boland B, Isbell M, Morin D, Shultz M, Baldwin R, Chan K, Karlsson A, Lin C, Taff A, West J, Fanucchi M, Van Winkle L and Plopper C. Naphthalene-induced respiratory tract toxicity: metabolic mechanisms of toxicity. Drug Metab Rev 2002; 34: 791-820.
[47] Politano VT, Lewis EM, Hoberman AM, Christian MS, Diener RM and Api AM. Evaluation of the developmental toxicity of methyl dihydrojasmonate (MDJ) in rats. Int J Toxicol 2008; 27: 295-300.
[48] Scognamiglio J, Jones L, Letizia CS and Api AM. Fragrance material review on methyl dihydrojasmonate. Food Chem Toxicol 2012; 50 Suppl 3: S562-571.
[49] Rachoin JaD, R. P. Infectious diseases and sepsis: recommendations for sepsis management. In: Sawroop SPaSaM, editors. Clinical Management of Shock: The Science and Art of Physiological Restoration. London, United Kingdom: IntechOpen; 2019. pp. 534-539. e532.