2019-01-01

Application Of Urinary Metabolites For Cancer Detection

Qin Gao
University of Texas at El Paso, qingaocn@live.com

APPLICATION OF URINARY METABOLITES FOR CANCER DETECTION

Qin Gao

Doctoral Program in Chemistry

APPROVED:

Wen-Yee Lee, Ph.D., Chair

Marc B. Cox, Ph.D.

Chu Young Kim, Ph.D.

Xiaogang Su, Ph.D.

Charles Ambler, Ph.D.
Dean of the Graduate School

Dedication

To my

Parents

with love

APPLICATION OF URINARY METABOLITES FOR CANCER DETECTION

by

Qin Gao.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry and Biochemistry

THE UNIVERSITY OF TEXAS AT EL PASO

May 2019

Acknowledgements

In my five years of studying at the University of Texas at El Paso (UTEP), I am privileged that I met many wonderful people in different fields, who help me pursue my dream of being a researcher.

I would like to express my deepest gratitude to my mentor, Dr. Wen-Yee Lee, for introducing me to the research field of analytical chemistry and clinical cancer research, and for her guidance, encouragement, enduring patience and constant support. Her dedication and motivation to research will continue to inspire me. Moreover, her commitment to the professional development of her students is impressively remarkable. Dr. Lee is always pleased to provide opportunities for her students to give presentations in all kinds of conferences, and provide strong recommendation letters whenever she is needed. Her helpful guidance equips me with competency in research, improve my scientific writing skills, and produce notable scientific publications and patents. I also really appreciate Dr. Lee's concern, encouragement, and plan for my professional development.

I am grateful to all my committee members: Dr. Marc B. Cox, Dr. Chu Young Kim, and Dr. Xiaogang Su. With distinct scientific backgrounds, they have provided numerous helpful advice, insight, and suggestions with their kindness and valuable time. I especially want to thank Dr. Su, with whom we have ongoing collaborations. Dr. Su's strong support of statistics was essential for my research.

I appreciate the support of my colleagues in the Dr. Lee's Research Group. In particular, I would like to thank those who helped me: Angela Encerrado Manriquez, Alma L Loya, and Tianqi Xiang. Other colleagues are likewise very much appreciated for their readiness to help and for the nice working atmosphere they created.

I would like to thank all the professors and staffs from Department of Chemistry and Biochemistry of UTEP for all their hard work and dedication, providing me the knowledge and means to complete my degree. I would like to express special thanks to Dr. River and Dr. Cox for

providing the access to use the equipment for my experiments. I would also like to thank Dr. Heinric Williams from Geisinger Medical Center (PA) and Dr. Annabi from the Clinic Internal Medicine (El Paso, TX) for providing the human clinical samples for prostate cancer and renal cancer diagnosis.

I want to thank my dear parents, my grandparents, my uncles, my aunts, and all my dear friends for their true love, encouragement, and firm support. It was impossible for me to continue my graduate study without them.

The funding agencies that supported my dissertation research are very much appreciated. I would like to acknowledge the financial support of the Graduate School (Dodson Research Grant), College of Science (Wiemer Family Student Endowment for Excellence), and Dr. Keelung Hong Gift Fund for VOC Cancer Research Studies and Chemistry Graduate Research Fellowship .Travel grant support from the Graduate School, College of Science, Department of Chemistry and Biochemistry and Student Government Association of UTEP is also gratefully acknowledged.

Abstract

Prostate cancer (PCa) is the 3rd most common cause of male cancer mortality in the US. Early diagnosis and treatment of PCa will improve the quality of care and reduce mortality. The prostate specific antigen (PSA) is commonly used in the current PCa screening, but its limitation has resulted in an intense search for more reliable biomarkers. Studies showed that dogs could differentiate PCa patients from negative control by sniffing their urine. As the odor profiles are generated by volatile organic compounds (VOCs), the finding suggests that particular VOCs could be linked to PCa, PCa risk levels and other cancers. Therefore the overarching goal of study was to develop a reliable, quick and patient-friendly diagnostic method for PCa screening to replace the current PSA testing and avoid unnecessary biopsy. The objectives of this study were 1) to develop and validate a urinary VOCs based model for PCa diagnosis; 2) to study the performance of urinary VOCs in differentiating high/intermediate PCa patients from low risk PCa patients; 3) to investigate the urinary VOCs for the early screening of other urological cancer, such as clear cell renal cell carcinoma (ccRCC); and 4) to evaluate the specificity of urinary VOCs models in PCa and ccRCC detection.

In the study of the PCa urinary VOCs profile, a VOCs based PCa model was developed and validated. Urine samples from 55 and 53 biopsy proven PCa positive and negative patients respectively were obtained for diagnostic model development. Urinary metabolites were analyzed by Stir Bar Sorptive Extraction coupled with Gas Chromatography-Mass Spectrometry. A PCa diagnosis model was developed and validated using innovative statistical machine-learning techniques. The analysis resulted in 254 and 282 VOCs for their significant association ($p<0.05$) with either PCa positive or negative samples respectively. Regularized logistic regression analysis and the Firth method were then applied to predict PCa prevalence, resulting in a final model that contained 11 VOCs. Under cross-validation, the area under the receiver operating characteristic curve (AUC) for the final model was 0.92 (sensitivity: 0.96; specificity: 0.80). Further evaluation of the developed model using a testing cohort yielded an AUC of 0.86. As a

comparison, the PSA-based diagnosis model of the same cohort of patients only rendered an AUC of 0.54.

Then, since the performance of VOCs was proved to be able to strongly discriminate PCa patients from controls, we hypothesized that urinary VOCs could be used to develop urinary VOC PCa prognostic models for cancer risk assessment. PCa is a heterogeneous disease ranging from indolent to life threatening stages. Another VOCs based PCa risk assessment model was developed through comparing high/intermediate risk patients with low risk patients. Based on the D'Amico risk scores system, these PCa patients were divided into two groups: 55 in low-risk group (indolent PCa, GS = 6, PSA< 10) and 34 in high/intermediate-risk group (clinically significant PCa, GS = 6 and PSA $\geqslant$ 10, or GS > 6 with any PSA values). Urine samples from 89 men who presented for transrectal ultrasound guided prostate biopsy for an elevated PSA or abnormal digital rectal exam were collected. Using the Wilcoxon rank sum test, 23 VOCs were found to be positively related to high/intermediate−risk PCa and 44 VOCs negatively associated. Regularized logistic regression analysis and the Firth method were then applied to predict PCa risk level, resulting in a final model that contained 11 VOCs. Under cross-validation, the area under the receiver operating characteristic curve (AUC) for the final model was 0.86 (sensitivity: 0.85; specificity: 0.80), which indicates a highly promising discrimination power of urinary VOCs in PCa high/intermediate risk assessment.

To test the performance of VOCs in other genitourinary cancers, we investigated the urinary VOCs profile of clear cell renal cell carcinoma (ccRCC the main type of renal cell carcinoma, RCC). The fast and reliable early screen of RCC enables better outcome and predication in patients. However, there is no recommended screening tests for RCC available clinically. A total of 108 urine samples were obtained from 71 patients who were undergoing partial or radical nephrectomy and 37 patients ccRCC negative based the imaged renal mass to identify the specific VOCs in urine for ccRCC screening. The VOCs based ccRCC diagnostic model was developed through the logistic regression in training set (57 ccRCC vs 31 controls) and validated in another testing set group (14 RCC vs 6 controls). A total of 8,266 VOCs were found

in the urine samples of training set. Using Wilcoxon Rank Sum Test in screening their bivariate association with ccRCC, 79 VOCs were found to be related to urine samples of ccRCC patients while 91 VOCs corresponding to RCC negative controls with statistical significance at p =0.05. After further selection with l_1 regularization, 15 VOCs were included in the RCC diagnostic logistic model. On the basis of predicted probabilities from the model via cross-validation, the area under the receiver operating characteristic curve (AUC) was found to be 0.87 and the sensitivity and specificity were 93% and 77% respectively. The VOCs based RCC diagnostic model were then validated in testing group. The results showed a promising diagnostic power of this VOCs model in ccRCC screening with AUC of 0.81, with 86% sensitivity and 83% specificity respectively.

Moreover, we cross examined the performance of the two PCa and ccRCC VOCs models. Four cohorts from the previous studies were involved in this analysis. The four groups contained (1) 55 PCa positive patients, PCa (+), (2) 53 PCa negative patients, PCa (-), (3) 55 ccRCC positive patients, ccRCC (+), and (4) 31 RCC negative patients, ccRCC (-). For the PCa model was cross examined between PCa (+) and the rest of the three groups combined, and the result showed an AUC of 0.834 with confidence interval 0.779 to 0.889. Then, the ccRCC model was validated through comparing ccRCC (+) with the rest of the three groups combined. The validation rendered an AUC of 0.779 with confidence interval 0.707 to 0.851. The cross evaluation verified the discrimination power of those 11 VOCs based PCa model and 15 VOCs based ccRCC model even in more complicated patients.

The investigations among PCa and ccRCC demonstrate and validate the clinical utility of a non-invasive urinary VOCs based diagnostic model in PCa and ccRCC screening. The VOCs based diagnostic model has the substantial potency and clinical value in PCa and RCC screening, and the analytical method was fast and highly translatable.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 VOCs emitted from human body

Volatile organic compounds (VOCs), the majority being organic in nature, are continuously generated in human body and released through breath, blood, skin, urine and fecal samples (Figure 1. 1).[1-4] These VOCs carry information of the physiological and metabolic status of the individual.[5] Furthermore, the VOCs emitted from different areas of the human body vary with age, diet, sex, physiological status and possibly genetic background. Therefore, body odors containing VOCs can be considered as individual 'odor-fingerprints'.[6]

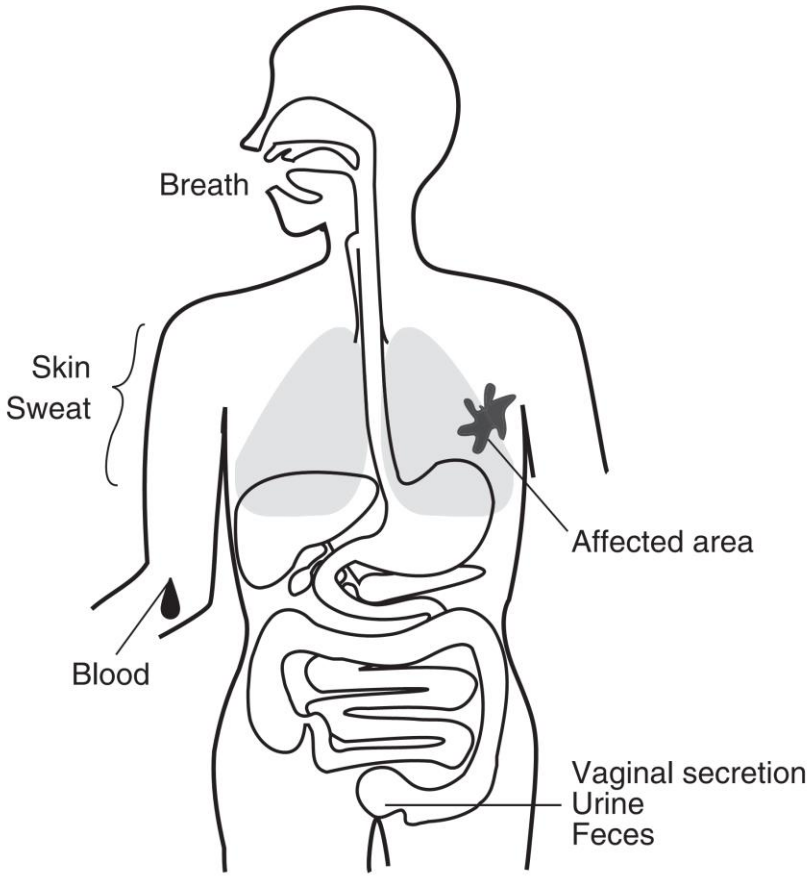


Figure 1. 1 Continuously generated from human body and released through breath, blood, skin, urine and feces. (*Source*: M. Shirasu and K. Touhara, 2011. Copyright © 2011 The Journal of Biochemistry.[6])

As VOCs are considered the metabolites of biological activities in human body, they exist in various parts of our system. Furthermore, pathological processes, occurring as a consequence of disease, can alter the production of specific VOCs which the body generates differently during

normal biological processes. In recent studies have demonstrated that sniffer dogs can differentiate cancer patients from control negative by sniffing their biological samples, such as urine.[7-10] Therefore, we can use VOCs as a predictive biomarker for disease detection.

A) VOCs in Blood

The specific VOCs in the blood have been reported to be useful in predicting and diagnosing diseases, such as ovarian cancer, colorectal cancer, lung cancer, and hepatic encephalopathy.[11-14] In the study of Horvath *et al*.[11], the trained dogs can differentiate ovarian cancer patients from the patients with other gynecological cancers and healthy control subjects through sniffing the blood samples from patients. And Wang *et al*.[12] carried out a study to identify the blood volatile compounds as biomarkers for colorectal cancer by collecting blood samples from 16 colorectal cancer patients and 20 healthy controls. Four metabolic biomarkers were found at significantly higher or lower level cancer patients. Further studies are needed to evaluate these results and to apply these findings to clinical diagnoses. Blood directly reflects the internal environment of the body, including nutritional, metabolic and immune status, which highly values the blood samples in disease-specific VOCs studies. However, obtaining blood samples is invasive and not easy due to the onus on patients, and pre-treatment of blood samples is also time-consuming. All of these factors have limited the use of VOCs in blood for diagnostic tool development.

B) VOCs in Breath

Exhaled breath contains VOCs that can be attributed to either exogenous or endogenous volatiles[15,16]. Endogenous volatiles consist of blood-borne compounds released to the environment via the lungs and/or compounds made from all classes of symbiotic bacteria. Numerous studies were conducted to investigate the potential of VOCs in breath in diseases diagnosis, especially lung cancer. Collecting breath samples is relatively simple, painless and non-invasive as compared to sampling blood. Phillips *et al.*[17] collected breath samples from 108 patients and a combination

of 22 VOCs in breath samples distinguished between patients with and without lung cancer. In the study of Peng et al.[18], an array of sensors based on gold nanoparticles were shown to be able to rapidly distinguish the breath of lung cancer patients from the breath of healthy individuals in an atmosphere of high humidity by training and optimizing sensors with the VOCs identified through gas chromatography/mass spectrometry (GC-MS). However, it is always challenging to distinguish contaminant environmental exogenous compounds from endogenously produced VOCs. Exogenous volatiles include compounds inhaled from the external environment, such as compounds produced following the oral ingestion of food and compounds derived from smoking cigarettes.

C) VOCs in Urine

The VOCs in urine are intermediate products or end products of a number of metabolic pathways, which may contain a variety of structural motifs with a particular odor, such as ketone, alcohol, furan, pyrrole and sulphide.[6] In some cases, characteristic urine VOCs profile have been directly linked to particular metabolic disorders. Some studies have been conducted to illustrate urinary VOCs profiles in infectious diseases[19,20] and different types of cancers, including prostate cancer (PCa)[21], renal cancer (RCa)[22] and bladder cancer (BCa)[23].

Urinary VOC patterns in cancer patients are often different from the patterns in urine samples from control subjects, although the differences depend on cancer types and even cancer stages. Khalid *et al.*[21] showed that urinary VOCs profile of prostate cancer patients can be discriminated from cancer free controls by using four VOCs, 2,6-dimethyl-7-octen-2-ol, pentanal, 3-octanone, and 2-octanone with accuracy as high as 71%. In an analysis of volatile human urinary metabolome for renal cell carcinoma (RCC), Monteiro *et al.*[22] reported a new and simple analytical approach developed to study the volatile human urinary metabolome and the results clearly showed the potential of volatile urinary metabolome to discriminate between RCC and control patients with 60.33% of the variability in principal component analysis (PCA). And according to Weber *et al.*[23], the best diagnostic performance they obtained through the comparison between healthy

volunteers and bladder cancer patients was 70% overall accuracy using a gas sensor array and pattern recognition.

These studies have proved the potential in searching for volatile diagnostic biomarkers in the urine of cancer patients. Due to the complexity of urine components, such as metabolites from ingested foods and drinks, and considerable variation among individuals, caution must be taken when determining whether or not a candidate VOC biomarker results from disease-related changes in metabolism and advanced computer processing of chromatographic data should be involved in identifying the VOC patterns.

D) Others

VOCs can also be continuously emitted from skin surface as sweat. Sweat is one of the less employed bio-fluids for discovery of markers. In the research conducted by Calderón-Santiago *et al.*[24], human sweat was collected and used as clinical sample to develop a screening tool for lung cancer. The capability of the metabolites identified in sweat to discriminate between patients with lung cancer versus smokers as control individuals was studied. The five metabolites identified in this study provided 80 % specificity and 79 % sensitivity. Mi-Jung *et al*.[25] also applied the analysis of sweat volatile organic compounds in forensic science. Although some of these VOCs result from internal hormonal or metabolic changes, many VOCs appear to be derived from symbiotic bacteria that live on the skin surface and metabolize and transform secreted compounds in sweat and sebum. Any alteration in homeostatic balance due, for instance, to some inherited metabolic disorder or bacterial infection of the diseased area can induce changes in both the quality and quantity of VOCs. For example, some infectious diseases or cancerous wounds develop characteristic and offensive odors.[6] Meanwhile, the contamination from the environment must also be taken into consideration, including the interference from the ambient air, humidity and cosmetics.

Human fecal samples represent dietary end-products resulting from digestive and excretory processes and intestinal bacterial metabolism. The investigation of fecal VOCs may reveal

potential health consequences and be the best non-invasive way of diagnosing gastrointestinal diseases. Distinct patterns of VOCs have been associated with fecal samples from patients with some types of bacterial infection, such as Vibrio cholera, , Clostridium difficile or Campylobacter jejuni infections.[26,27]. Batty *et al.*[28] reported the use of volatile fecal metabolome in screening for colorectal cancer with 78% specificity and 72% sensitivity. VOCs may also be contained in other types of bio-fluids, such as vaginal secretions, which accurately reflect the stages of menstrual cycles

In summary, VOCs can be emitted from different types of biological samples of human body and carry "odor fingerprint" of the individual (Table 1. 1). Pathological processes, such as infection and endogenous metabolic disorders, can influence our daily odor fingerprints by producing new VOCs or by changing the ratio of VOCs that are produced normally.[6] And that is why these VOCs could potentially be attractive to physicians, physiologist, and surgeons as potential aids to clinical diagnosis and therapeutic monitoring. However, those VOCs may be affected by various factors, such as age, sex, drug therapy, diet and smoking. Therefore, care must be taken when investigating disease samples and when comparing them to control samples.

Table 1. 1 VOCs, as "odor fingerprint", could be emitted from different types of biological samples of human body.

The origins of odor	Published Paper	Disease Detection	Method	Sample Size	Study Results	
					# of VOCs	How reliable?
Blood	Horvath *et al.*[11]	Ovarian cancer	Trained dogs	N/A	N/A	Tissue tests: 100% sensitivity and 95% specificity. blood tests: 100% sensitivity and 98% specificity
	Wang *et al.*[12]	Colorectal cancer	SPME-GC-MS	16 cancer patients and 20 healthy controls	4	Lower level VOCs ($p<0.01$): Higher level VOCs ($p<0.05$):
Breath	Phillips *et al.*[17]	Lung cancer	GC-MS	60 cancer patients and 48 non-cancer controls	22	100% sensitivity and 81.3% specificity
	Peng et al.[18]	Lung cancer	Sensors based on gold nanoparticles	N/A	42	Accuracy >86%
Urine	Khalid *et al.*[21]	Prostate Cancer	SPME-GC-MS	59 cancer patients and 43 non-cancer controls	4	AUC 0.76 Accuracy as high as 74%
	Monteiro *et al.*[22]	Renal cell carcinoma	GC-MS	N/A	N/A	N/A
	Weber *et al.*[23]	bladder cancer	Gas sensor	30 cancer patients and 59 non-cancer controls	N/A	70% overall accuracy; 70% sensitivity and 70% specificity
Sweat	Calderón-Santiago *et al.*[24]	Lung cancer	LC-MS	41 cancer patients and 55 non-cancer controls	16	specificity/sensitivity pair (80 and 79 %
Feces	Batty *et al.*[28]	Colorectal cancer	Ion flow tube mass spectrometry (SIFT-MS)	31 high risk patients and 31 low risk or non-cancer controls	N/A	Accuracy 75% with 78% specificity and 72% sensitivity

1.2 Extraction and Detection of VOCs as potential method for disease diagnosis

1) VOCs Extraction

Since the low concentration of VOCs present in biological samples of human body, the extraction and pre-concentration are crucial for the analysis of VOCs of interest[29] and may affect the reliability and accuracy of the analysis[30].

To increase the reproducibility, selectivity, and extraction capacity and simplify the sample preparation steps, solid-phase mirco-extraction (SPME) was developed to facilitate the rapid sample preparation[31]. SPME is an extraction technique (

Figure 1. 2), in which a fine bare fused silica fiber or a fine silica fiber coated with a thin layer of a selective coating (can be solid or liquid) is used to extract organic compounds directly from aqueous samples for instrumental analysis by GC or GC-MS[32]. There are two types of extraction based on the different samples: 1) the direct immersion SPME extracts VOCs by immersing the fiber in liquid samples, and 2) the headspace SPME by suspending fiber in headspace above the liquid phase. The analytes are firstly adsorbed during extraction on the surface of the fiber materials as a result of chemical bonding, and then absorbed as being dissolved into coating materials[33]. Four types of polymers are widely used as the coating materials, polydimethylsiloxane (PDMS), divinylbenzene (DVB), polyacrylate (PA), and polyethyleneglycol (PEG). Those materials could also be used through the combination blended with carboxen (CAR)[34]. After pre-concentrating, the fiber with analytes trapped on its coating materials is injected to the instruments and release analytes through thermal desorption. Deng *et al.*[13] developed a simple, rapid and sensitive SPME/GC-MS method for the investigation of volatile biomarkers in lung cancer blood. Poli *et al.*[30] evaluated the potential of aldehydes from exhaled breath in the diagnosis of non-small cell lung cancer by means of SPME/GC-MS with 93% accuracy and precision between 7.2-15.1%. Monteiro *et al.*[22] studied the volatile human urinary metabolome difference of RCC and healthy individuals through the headspace SPME sampling

coupled with gas chromatography-ion trap/mass spectrometry (GC-IT/MS). Wang *et al.*[12] analyzed the VOCs in the blood samples from colorectal cancer patients and healthy controls with headspace SPME sampling. Khalid *et al.*[21] also applied SPME in headspace of urine samples to identify the specific urinary VOCs for the detection of prostate cancer.

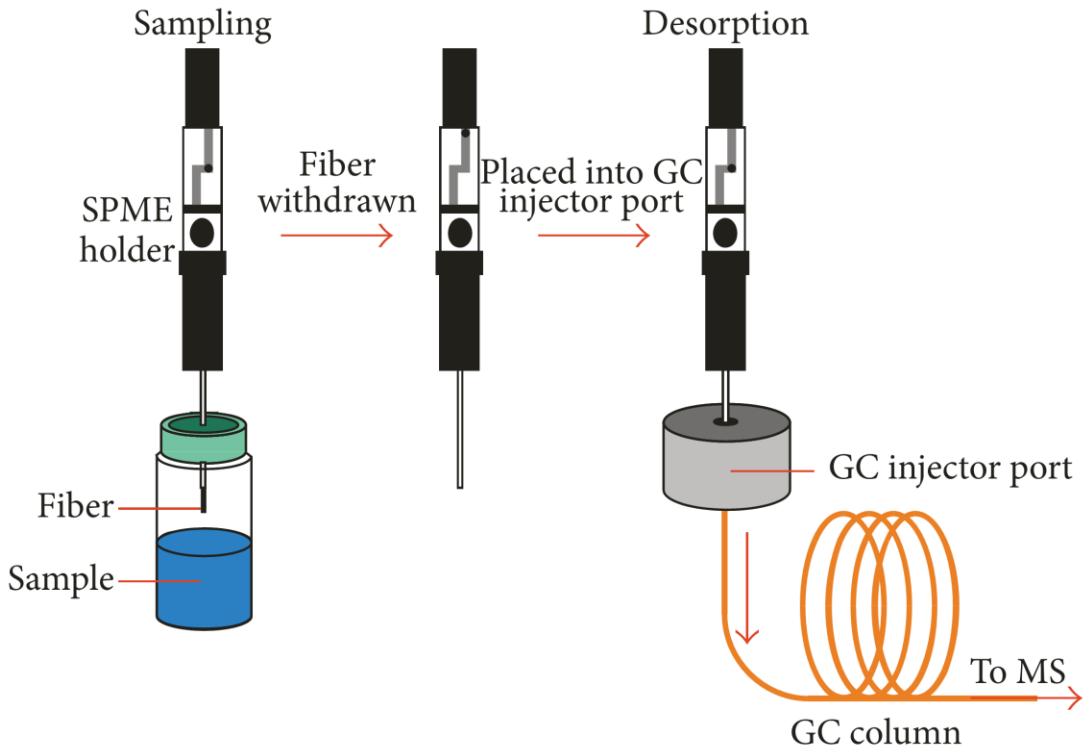


Figure 1. 2 The use of solid phase micro-extraction-gas chromatography-mass spectrometry (SPME-GC-MS) (Source: Kamila Schmidt and Ian Podmore, 2015. Copyright © 2015 The Journal of Biomarkers.[29])

Similar to the theory of SPME, another novel approach for sample enrichment is referred as stir bar sorptive extraction (SBSE), reported by Baltussen *et al.*[35]. Similar to SPME, SBSE uses stir bars coated with the sorbent PDMS. The results of experiments conducted Baltussen *et al.* indicated that the stir bars coated with PDMS present higher efficiency than SPME in the pre-concentration of analytes from aqueous samples, with up to a 500-fold increase in sensitivity when stirring between 30 to 60 min. The high efficiency could be contributed to the increased amount

of PDMS coated on the stir bars. Furthermore, the volatile compounds can also be easily and conveniently handled due to the absence of drying step. Therefore, SBSE can be applied in the analysis of VOCs in different types of aqueous samples, as well as the biological fluids. Melo *et al.*[36] carried out an analysis of antidepressants in plasma samples using SBSE and liquid chromatography (LC) with high extraction efficiency. Soini *et al.*[37] proved the high reproducibility of SBSE in quantitative comparisons of the urinary profiles with relative standard deviations (RSD) of 1-5% for a wide range of compounds. In one of our studies, SBSE was successfully applied in identifying the specific urinary volatile organic compounds for the diagnosis of prostate cancer.[38] In addition, the coated stir bar could also be used as headspace sorptive extraction (HSSE).[39,40]

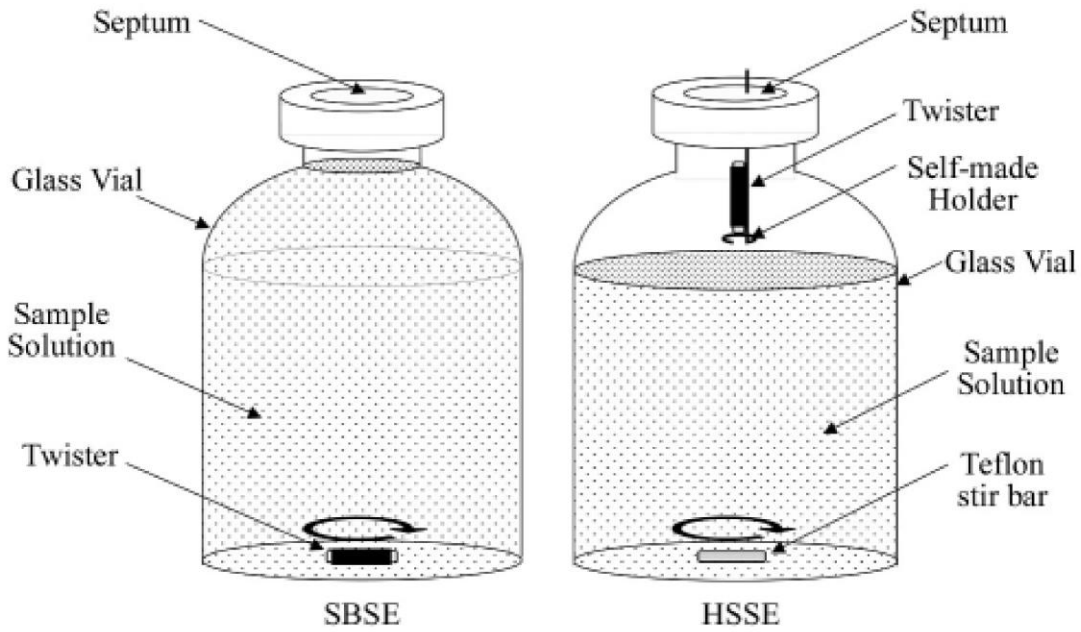


Figure 1. 3 The set-ups of SBSE and HSSE. (Source: Ochiai *et al.*, 2001. Copyright © 2001 The Royal Society of Chemistry.[40])

Other less common extraction techniques, such as purge and trap[41], single drop micro-extraction[42], etc. were also applied in VOCs extraction.

2) VOCs Detection

VOCs are easy to be detected by using analytic instruments, like GC-MS, proton transfer reaction-mass spectrometry (PTR-MS), selected ion flow tube-mass spectrometry (SIFTMS) or gas sensors.[21,29,43-45]

As one of the most commonly used analytical technique, GC-MS is widely used in the investigation of potential VOC cancer biomarkers based on its sensitivity and reliability in analyte identification.[12,13,22,30,38] It was shown that GC-MS can reach the ppb and low ppt levels in VOC analysis with the pre-concentration steps.[46,47] Fuchs *et al*.[46] analyzed the breath gas aldehydes from lung cancer patients using GC-MS. The concentrations in their study ranged from 7 pmol/l (161 pptV) for butanal to 71 nmol/l (1,582 ppbV) for formaldehyde. In another study using GC-MS conducted by Ligor *et al.*[47], the limit of detection was in the range of 0.05 to 15.00 ppb. Meanwhile, it provides the most detailed information of VOCs profiles and identifies analytes with most certainty. However, GC-MS instruments are often expensive. Compared to GC-MS, PTR-MS and SIFT-MS do not require a pre-concentration step and can work in real time, which make these two better instant quantification techniques for VOCs analysis.[48,49] Wehinger *et al*.[48] identified ions at m/z 31 and 43 from the exhaled breath that can best discriminate the primary lung cancer patients and controls using PTR-MS. As mentioned in this study, even though the technique is simple and time-saving for larger clinical evaluation, it is not possible for PTR-MS to differentiate between compounds with the same molecular mass.

Some other detection techniques are also used in the analysis VOCs emitted from human, such as ion mobility spectrometry (IMS)[50]. Compared with GC-MS, IMS gives a tenfold higher detection rate of VOCs (500 seconds for IMS vs. 1h for GC-MS per sample). In one study of detecting VOCs in exhaled breath of patients with lung cancer, the IMS was used by Westhoff *et al.*[50] and a combination of 23 peak regions were identified to discriminate the cancer patients and controls without error.

In addition, several types of electronic noses have been used in the studies of VOCs in cancer.[18,23,51] Natale *et al.*[51] investigated the possibility of using electronic nose to identify the lung cancer patients from controls. The results in their study indicated a 100% of classification of lung cancer affected patients and 94% of controls. These sensors used in this study showed a good sensitivity towards the compounds identified previously as potential lung cancer markers. However, electronic noses are designed to recognize the VOCs found in established studies but not to identify any unknown VOC patterns. Compared to the mass spectrometry based techniques, the electronic nose is less time consuming and enables the potential of cheap, rapid, simple, and miniature detection devices[52,53]. However, electronic noses are sensitive to moisture, less sensitive, and with poor reproducibility.[54,55] Additionally, electronic noses can only allow the semi-quantitative detection of VOCs[56].

1.3 VOCs and Urological Cancers

Cancer is a leading cause of death and disability globally, impacting more than 14 million people each year.[57] Urological cancers, such as prostate cancer (PCa), renal cancer (RCa), and bladder cancer (BCa), are a major cause of morbidity and mortality worldwide.[58] In 2018, about 164,690 new cases of PCa, 65,340 of RCa, 81,190 of BCa and about 29,430 deaths in PCa, 14,970 in RCa, 17,240 in BCa are estimated in United States according to the American Cancer Society.[59] In the United States, PCa is the most common cancer and the third leading cause of death in men. About 1 in 7 men will be diagnosed with PCa during his lifetime.[59] RCa and BCa also account for more than 2% and 4% of cancer mortality in the United States.[59]

Diagnosis and treatment for each these urological cancers is associated with different, but overlapping, clinical challenges.[58] High-throughput genomic screening, proteomic profiling, and metabolomics analysis of related functional protein molecules provide a large amount of informational data and overview of clinical changes of cancer development and progression. The cells, proteins, and metabolites in urine originated from kidney, prostate, and bladder could provide

research information for biomarkers searching, such as genomics, proteomics, and metabolomics.[60-62] Urine, as a source of excretion from the urological system, is an ideal body fluid for the investigation and detection of biomarkers for those urological cancers. Moreover, urine can be easily sampled and non-invasive.

As early diagnosis and treatment of those urological cancers will improve the quality of care and reduce mortality, there is a high demand of reliable, quick and patient-friendly diagnostic method for cancers screening. As aforementioned, several studies have demonstrated that sniffer dogs can differentiate cancer patients from controls by sniffing their urine.[7-10] Cornu *et al.*[8] reported the trained dog detected PCa by smelling urine with 91% for both sensitivity and specificity. The study from Willis *et al.*[10] provided further evidence that volatile compounds found in urine can be identified by trained dogs with 73% sensitivity and 92% specificity. Additionally, VOCs are easy to be detected by using analytic instruments, like gas chromatography-mass spectrometry (GC-MS), or further developed gas sensors.[21,43-45] All of those proved VOCs, particularly in urine, as a desirable disease markers for its non-invasiveness, non-injury, easy detection, high sensitivity and high specificity. As one of the most promising metabolomics approaches in cancer detection, the analysis of VOCs can potentially serve as a safe, non-invasive, and specific test for the early detection of those urological cancers.

1.3.1 VOCs in Prostate Cancer

Currently, PCa are screened by the prostate-specific antigen (PSA) blood test and/or the digital rectal exam (DRE). If the PCa is suspected based on the results of screening tests or other symptoms, further tests, including trans-rectal ultrasound (TRUS) prostate biopsy will be required to confirm the diagnosis.[63] Furthermore, techniques used in advanced stages, such as bone scans, computed tomography (CT) scan, and magnetic resonance imaging (MRI), may involve X-rays, magnetic fields, sound waves and radioactive substances which can lead to the second injury of

cancer patients.[63] Diagnostic methods which can reduce stress and be more patient-friendly are needed.

Currently, the gold standard for PCa screening is PSA. Adding to the fact that PSA is not cancer specific, there is no reliable PSA threshold that can accurately distinguish men with or without cancer.[64] Therefore, anyone suspected of PCa must undergo prostate biopsy. Depending on the PSA level, up to 80% of men were found no cancer based on their biopsy results.[65] Therefore, there is a significant interest in finding a more accurate PCa-specific biomarker. Khalid *et al.*[21] showed the discrimination power of urinary VOCs profile in differentiating PCa patients from controls with 71% accuracy based on only 4 VOCs. In one of our studies[38], the performance of VOCs has been tested and validated with AUC 0.92 (96% sensitivity and 80% specificity). The VOCs based prostate cancer diagnosis tool would be promising in future clinical use.

Several biomarkers have been developed to improve upon the limitations of serum PSA including Iso-PSA, prostate cancer antigen 3 (PCA3), 4Kscore, Prostate Health Index (PHI), TMPRSS2:ERG and ConfirmMDx.[66-71] Among those markers, IsoPSA, PHI and 4Kscore are all PSA-based assay for PCa risk assessment.[66,69,71] PCA3 is a noncoding RNA that is prostate specific and highly overexpressed in prostate cancer.[72] TMPRSS2-ERG gene fusions are reported to be the predominant molecular subtype of prostate cancer.[73] ConfimMDx is an epigenetic test for PCa diagnosis before prostate biopsy.[70] Again, these methods are not able to provide satisfactory screening capability for PCa.

1.3.2 VOCs in Renal Cancer

The most common type of kidney cancer is renal cell carcinoma (RCC) consisting about 90% of kidney cancer cases. RCC is a heterogeneous malignancy, both morphologically and genetically, which is classified into different histologic subtypes, including clear cell RCC (most common one), papillary RCC, chromophobe RCC and other less common subtypes.[74-76] The outcome of RCC is usually unpredictable even after a long period of asymptomatically

development and progression.[77] Therefore, its diagnosis is often incidental through the use of medical imagology and is frequently detected at an advanced stage and metastatic when detected clinically.[78] Additionally, RCC is particularly challenging to treat because of its relative insensitivity to radiotherapy and conventional chemotherapy drugs.[79] The early screening of RCC could improve the outcome of diagnosis. However, no early screening method is recommended to screen for kidney cancer clinically in people at average risk or increasing risk.

The potential of urinary VOCs used in RCC diagnose has been highlighted in previous studies.[22,80,81] The purpose of most previous studies were focused on the searching of specific VOCs in RCC patients without further validation.[22,80] In the study reported by Marica Monteiro in 2017[81], the selected VOCs was validated in different patients group besides the searching of specific VOCs, but the performance of VOCs in differentiating RCC patients and controls was not determined. Besides, two urinary exosomal proteins, AQP-1 and PLIN2 have shown promise as the biomarkers in RCC diagnosis.[82] It should be noted that AQP-1 and PLIN2 can be found in clear cell and papillary RCC but not in the chromophobe subtype of RCC. However, VOCs based screening has great potential to be developed as a more universal screening tool of almost all types of RCC or even specific screening tool for each type of RCC because of the metabolic distinction shown with each selected VOC between cancer patients and controls, Unlike the ELISA detection methods of AQP-1 and PLIN2, the VOCs based diagnostic model could be developed as a high throughput and fast screening method in clinic enabled by high performance GC/MS and statistic assistance.

1.3.3 VOCs in Bladder Cancer

Bladder cancer (BCa) is the second most common genitourinary malignant disease in United States.[83] And it is also a heterogeneous malignancy, with different histologic subtypes, including transitional cell carcinomas (90%), squamous cell carcinomas (5%), and adenocarcinomas (less than 2%).[84] The most common symptom of BCa (in 80%-90% of the

patients)[85] is hematuria, or blood in the urine, and others including complaints of dysuria (painful urination), increased frequency or urgency of urination, failed attempts to urinate, a mass in the bladder or a ureteral obstruction.[86,87] Intravenous pyelography, cystoscopy, transurethral biopsy, and imaging techniques, such as magnetic resonance imaging and computerized tomography scan, are always involved in the clinical diagnosis procedures of potential BCa.[87,88] On the other hand, no early screening method for bladder cancer is recommended in United States. [89]

Many studies are attempting to identify genetic and chemical markers in order to complement the use of clinical features and better assess the risk level of BCa.[90,91] The overexpression of the *p53* gene, cells containing multiple aneuploid cell lines, and the expression of the Lewis- x blood group antigen were found to be the markers of high risk BCa.[90] Furthermore, nuclear matrix protein 22 (NMP22) and bladder tumour antigen (BTAstat) are more sensitive (50–85% and 50–70%), but less specific (60-70%), than urine cytology, which have been approved by FDA as protein markers of bladder cancer.[23,92]

Recently, VOCs are also suggested in different studies that have potential in differentiating patients of BCa from controls. And according to Weber *et al*.[23], the best diagnostic performance they obtained through the comparison between healthy volunteers and bladder cancer patients was 70% overall accuracy (70% sensitivity and 70% specificity) using a gas sensor array and pattern recognition. The results of another study using gas sensors, reported from Khalid *et al.*[93], also showed potential of VOCs for the diagnosis of bladder cancer (the best performance: 100% sensitivity and 94.6% specificity). All those studies have revealed the potential of VOCs used in bladder cancer diagnosis.

1.4 Potential molecules pathways related to VOCs profile for cancer direction

1.4.1 Androgen signaling and one-carbon metabolism

The androgen receptor (AR), is a nuclear receptor that is activated by binding either of the androgenic hormones, testosterone, or dihydrotestosterone in the cytoplasm and then translocating

into the nucleus.[94,95] It plays an essential and important role in PCa initiation, progression, and metabolic adaptation that takes place during PCa progression. As a transcription factor, the AR directly affects essential catabolic and biosynthetic pathways through modulating the expression of related effectors and regulators. On the other hand, the AR, as a modulator of the one-carbon metabolism, can also affect epigenetic processes, DNA metabolism, and redox balance indirectly, which are all important factors in tumorigenesis.[96]

One-carbon metabolism involves a complex network with two central cycles: (1) the folate cycle and (2) the methionine cycle. One-carbon metabolism also regulates essential processes including DNA synthesis and repair, epigenetic methylation reactions, redox homeostasis, and protein synthesis. The balanced flux through these four pathways, e.g. folate cycle, methionine cycle, transsulfuration pathway, and polyamine synthesis, is essential for cellular homeostasis. (Figure 1. 4 Part A),[96] and disruptions of the balanced flux could contribute to the pathogenesis of many diseases, including cancer.[97] Cancer creates a demand and dependency on one-carbon metabolism. For example, methyl group availability for methyltransferases that modulate gene expression via epigenetic mechanisms is influenced by flux within the folate cycle and methionine cycles.[98,99] Alterations in one-carbon metabolism may contribute to tumorigenesis through fueling DNA synthesis, changing the DNA and histone methylomes, promoting protein translation, driving cell cycle progression, and modulating redox balance. These changes can in turn promote sustained proliferation, induce tumorigenic gene expression changes, contribute to genomic instability, and promote survival—all important processes in tumorigenesis and cancer progression.[96]

The progression and metastasis of tumors were associated with metabolite increases in glutathione and cysteine/methionine metabolism pathways. For example, clear cell RCC is characterized by broad shifts in central carbon metabolism, one-carbon metabolism, and antioxidant response, reported by Hakimi *et al.*.[100] Bridging the gap between the Cancer Genome Atlas (TCGA) transcriptomic profiling and the metabolomic data in their studies, the authors were

able to integrate the pathway-level metabolic atlas and to demonstrate discordance between transcriptome and metabolome.

Studies in PCa cell lines demonstrate AR-regulation of one-carbon metabolism enzymes, and altered cellular methylation potential in response to androgens.[101-104] For example, sarcosine, a methylated metabolite of the one-carbon pathway, was found be accumulated in PCa clinical samples.[102] In the prostate, androgens and the AR regulate the activity and/or expression of several enzymes involved in the one-carbon metabolism pathways, specifically enzymes involved in S-adenosyl-methionine (SAM) homeostasis and the entry into the transsulfuration and polyamine synthesis pathways (Figure 1. 4 Part A). Studies directed to identify AR transcriptional networks in different models of PCa have demonstrated an involvement of the AR in global metabolism by directly targeting enzymes involved in several metabolic processes.[105-108] These findings illustrate the role of the AR in PCa tumorigenesis by controlling metabolism, and the value of integrating metabolomic profiling and gene expression analysis for the identification of new biomarkers and therapeutic targets. Also, these observations emphasize the link between the AR and one-carbon metabolism, and the potential effects that changes in AR signaling, that can occur with disease progression, may have on essential cellular processes.

1.4.2 The effect of Metabolic phenotype on fatty acid and phospholipid synthesis

The zinc accumulating and citrate synthesizing phenotype is the hallmark of the healthy prostate epithelial cell.[109,110] However, PCa cells reverse this phenotype and adopt a zinc wasting, citrate oxidizing phenotype, thereby representing a major shift in energy metabolism.[111] This shift allows these cells to utilize the Krebs cycle and subsequent oxidative phosphorylation (Figure 1. 4 Part C). It has long been identified that PCa cells do not conform to the standard Warburg effect seen in most cancers, which described in the early to mid-1900s by Otto Warburg.[112] Malignant cells shift their dominant ATP producing pathway away from oxidative phosphorylation to aerobic glycolysis.[112] Unlike most cancer cells that resort to aerobic glycolysis, prostate cancer cells exhibit a higher level of citric acid cycle activity compared to benign cells.[110] The increased

activity of citric acid cycle, essential for the progression of malignancy, was induced by the inability of malignant prostate cells to accumulate high zinc levels, which inhibits citrate oxidation.[113]

Another metabolic hallmark of many cancer cells is the enhanced lipogenesis, arising from increased activities of fatty acid biosynthetic enzymes.[114-116] Clear cell RCC (ccRCC) is histologically defined by its lipid and glycogen-rich cytoplasmic deposits.[117,118] In the study of Du *et al.*[118], the lipid deposition of ccRCC was investigated with focus on the carnitine palmitoyltransferase 1A (*CPT1A*), as a direct HIF target gene. Prostate cancer cells often utilize lipids derived from androgens through the expression of the AR.[119] However, these cells can also utilize *de novo* lipid synthesis to produce fatty acids in order to obtain energy. This shift to a lipid-producing phenotype is a key turning point in the progression of prostate cancer. The *de novo* lipid producers have ability to produce the key energetic molecules for growth without the regulation of androgens (Figure 1. 4 Part C).[120] Clinically, this is problematic as it represents a disease that is unresponsive to androgen deprivation therapy, known as castration-resistant prostate cancer.[121]. These producers include fatty acid synthase (FASN), sterol regulatory element binding protein 1 (SREBP1), and steroyl CoA desaturase. Among them, the enzyme FASN functions to help synthesize long-chain fatty acids. It is believed that unregulated FASN activity within prostate tissue is the beginning of malignant phenotype, and has been argued to be necessary for PCa growth maintenance.[122] The use of lipid by the PCa cells illustrates that these cells bypass potential degenerative pathways, and rather utilize the anabolic pathways in order to maintain energy and growth.[123] A variety of fatty acid moiety were detected in our preliminary study and that supports the importance of specific VOCs in PCa.

Additionally, phospholipids, also as the downstream products of enhanced lipogenesis, in the cancer cell membrane have been found to be abnormal compared with normal cells. The plasma membrane of normal cells is characterized by an asymmetric distribution of various phospholipids over two membrane leaflet. Phosphatidylethanolamine (PE) resides in the inner leaflet facing the cytosol (Figure 1. 4 Part C). The disrupted membrane asymmetry of cancer cell exposes PE to

extracellular space, which can serve as a molecular target for anticancer therapy.[124] The increasing need of PE in cancer cell may correlate the excessive consumption of ethanolamine (mostly appear in control group of our preliminary study) and enhanced lipogenesis.

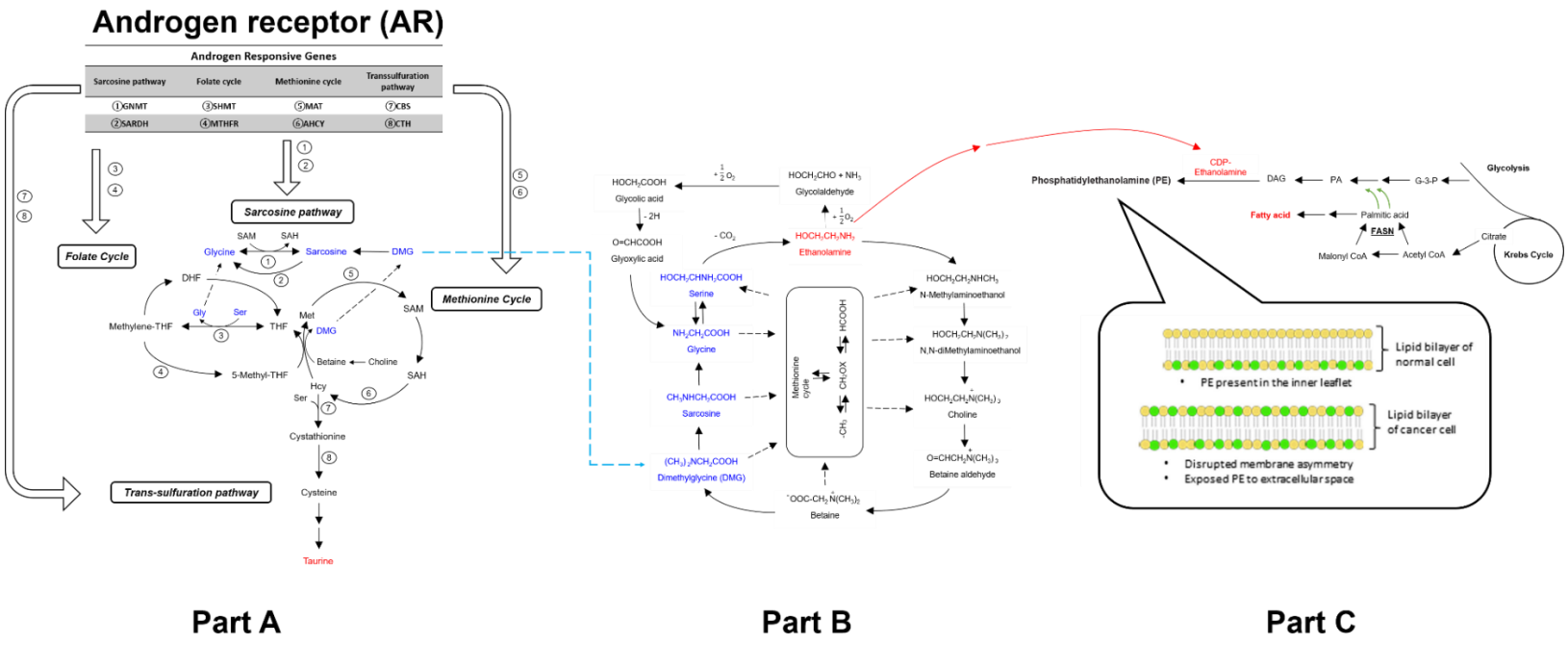


Figure 1. 4 A comprehensive illustration of Androgen signaling, one-carbon metabolism, and metabolic phenotype. Part A. One-carbon metabolism involves a complex network with four pathways: (1) folate cycle; (2) methionine cycle; (3) transsulfuration pathway; (4) sarcosine pathway. In the prostate, androgens and the AR regulate the activity/expression of several enzymes involved in the one-carbon metabolism pathway. Enzyme abbreviations are as follows: SARDH: Sarcosine Dehydrogenase; SHMT: Serine hydroxymethyltransferase; GNMT: Glycine-N-methyltransferase; MTHFR: Methylene tetrahydrofolate reductase; MAT: Methionine adenosyltransferase; AHCY: S-adenosylhomocysteine hydrolase; CBS: Cystathionine beta-synthase; CTH: cystathione gamma-lyase or gamma-cystathionase. **Part B**. Hypothetical cycle of metabolism involving glycine, serine, ethanolamine, choline, and betaine.[101] **Part C**. Enhanced lipogenesis, arising from increased activities of fatty acid biosynthetic enzymes (including ACC1, FASN, and stearoyl CoA desaturase (SCD1)), is a metabolic hallmark of many cancer cells.[106-108] In addition, the plasma membrane of normal cells is characterized by an asymmetric distribution of various phospholipids over two membrane leaflet. PE resides in the inner leaflet facing the cytosol. The disrupted membrane asymmetry of cancer cell exposes PE to extracellular space. Furthermore, PE is also highly exposed on endothelium cells in tumor vasculature. PA, phosphatidic acid; PC, phosphatidylcholine; DAG, diacylglycerol; CDP-ethanolamine, Cytidine diphosphate ethanolamine.

1.5 Research Objectives

The overarching goal of this research was to develop and validate the urinary VOCs based diagnostic tool for prostate cancer and renal cancer for early screening and even the risk stratification. Specific objectives of the project are described as follows.

1) Develop and validate a urinary VOCs based model for prostate cancer diagnosis.

Identify specific VOCs in urine that can differentiate prostate cancer patients from controls. The urinary VOCs based PCa diagnostic model would be validated through training-testing set. Evaluate the model with external cohort of patients to verify the performance of the VOCs based diagnostic model in differentiating PCa patients from controls.

2) Develop and validate a urinary VOCs based model for prostate cancer risk assessment.

Examine specific urinary VOCs in PCa risk assessment. The urinary VOCs profile were compared between PCa low risk group and PCa high/intermediate group. A risk predictive modeling using logistic regression and cross-validation were developed and assessed.

3) Apply the VOCs tool for other cancer early detection study. Develop and validate a urinary VOCs based model for clear cell renal cell carcinoma diagnosis.

Screen the specific urinary VOCs for ccRCC early detection. The urinary VOCs based ccRCC diagnostic model was validated through training-test set. The performance of those selected VOCs were evaluated again through the external cohort of ccRCC patients.

4) Testing the specificity of the two VOCs models by cross examining the discrimination power of VOCs in prostate cancer and clear cell renal carcinoma. The performance of developed urinary VOCs based PCa and ccRCC diagnostic models in cohorts of patients with increased complexity would be tested.

Chapter 2: Application of Urinary Volatile Organic Compounds (VOCs) for the diagnosis of Prostate Cancer

Reproduced in part with permission from the *Genitourinary Cancer*.

- Parts of this chapter were published in *Genitourinary Cancer* as a research article in 2019. (Qin Gao, Xiaogang Su, Michael Hani Annabi, Brielle R Schreiter, Thomas Prince, Andrew Ackerman, Sara Morgas, Valerie Mata, Heinric Williams,*, Wen-Yee Lee,*. *Genitourinary Cancer*, 2019)
- This chapter introduces a study to address the critical and unmet need for a simple, effective, and sensitive diagnostic tool for prostate cancer. We collected urine samples from 108 biopsy proven prostate cancer positive and negative patients to develop a metabolomics based model to detect prostate cancer. The model was validated and showed high discriminating power in prostate cancer detection.

Abstract:

Background: Prostate cancer (PCa) screening using serum prostate specific antigen (PSA) testing has caused unnecessary biopsies and over-diagnosis due to its low accuracy and reliability. Therefore, there is an increased interest in identifying better PCa biomarkers. Studies showed that trained dogs can discriminate PCa patients from unaffected men by sniffing urine. We hypothesized that urinary volatile organic compounds (VOCs) may be the source of that odor and could be used to develop urinary VOC PCa diagnosis models.

Methods: Urine samples from 55 and 53 biopsy proven PCa positive and negative patients respectively were initially obtained for diagnostic model development. Urinary metabolites were analyzed by Gas Chromatography-Mass Spectrometry. A PCa diagnosis model was developed and validated using innovative statistical machine-learning techniques. A second set of samples (53 PCa positive and 22 negative patients) were used to evaluate of were used to evaluate the previously developed PCa diagnosis model.

Results: The analysis resulted in 254 and 282 VOCs for their significant association ($p<0.05$) with either PCa positive or negative samples respectively. Regularized logistic regression analysis and the Firth method were then applied to predict PCa prevalence, resulting in a final model that contains 11 VOCs. Under cross-validation, the area under the receiver operating characteristic curve (AUC) for the final model was 0.92 (sensitivity: 0.96; specificity: 0.80). Further evaluation of the developed model using a testing cohort yielded an AUC of 0.86. As a comparison, the PSA-based diagnosis model only rendered an AUC of 0.54.

Conclusion: The study describes the development of a urinary VOC-based model for PCa detection.

Clinical Practice Points

- Due to the low specificity of current prostate cancer screening with serum Prostate-Specific Antigen (PSA), up to 80% of men who are found to have elevated PSA value and subsequently undergo prostate biopsies will either not be diagnosed with prostate cancer (PCa) or be diagnosed with an indolent form of PCa, which is unlikely to cause significant morbidity or mortality. It is urgent to address the problem of over-diagnosis and overtreatment in prostate cancer care.
- We developed urinary metabolites (i.e. VOCs) based model to detect PCa with a higher accuracy (87%) than PSA alone (59%) in PCa screening, thus potentially providing a better alternative than current PSA test to accurately and reliably diagnose prostate cancer.
- As a non-invasive and sensitive tool, the VOC model is ideally suited for early diagnosis of PCa. Future validation of our model will enable clinically relevant and cost-effective, point-of-care PCa screening and prognosis with the potential to save health care costs.

2.1 Introduction

In United States, prostate cancer (PCa) is the most common non-cutaneous cancer and the second leading cause of cancer death in men [59]. Serum prostate specific antigen (PSA) is the most commonly used biomarker for PCa. However, PSA lacks accuracy for PCa diagnosis and, as a result, contributes to unnecessary prostate biopsies, over-diagnosis and over-treatment of the disease [64]. Therefore, finding more sensitive and reliable PCa screening methods is greatly needed.

In United States, approximately 19 million men undergo serum Prostate-Specific Antigen (PSA) blood testing annually [125], resulting in 4.7 million abnormal PSA test results (≥4.0 ng/mL) leading to 1.3 million biopsy procedures [126] . Only 22% of men undergoing their first biopsy will be diagnosed with PCa [127]. In addition, more than a third of the men with PCa on prostate biopsy will have clinically indolent or low-risk disease. Thus, a better screening assay that detects clinically significant PCa could potentially eliminate nearly 80% (~1 million) prostate biopsies,

avoiding the unnecessary expense and the associated risks of pain, infection, bleeding, difficulty urinating and anxiety [128]. Therefore, developing an effective diagnostic test capable of not only identifying men with PCa, but more importantly, those with clinically significant disease is needed.

Metabolic reprogramming is a well-recognized hallmark of cancer and metabolomic approaches have shown promise in oncology diagnosis, prognostic and treatment [129]. Studies reported that trained animals can detect cancers by "sniffing" the urine from patients with breast [130], bladder [131], prostate [8,132], colorectal [133], and lung cancer [134]. In PCa studies, trained dogs were able to discriminate between patients with and without disease (sensitivity and specificity >90%) [8]. Since the odor of urine perceived by the dogs is produced by volatile organic compounds (VOCs), gas chromatography-mass spectrometry (GC/MS) has been proposed to detect VOCs in men for PCa.

In general, VOCs can be produced from the human body and released through breath, blood, skin, urine and feces [135]. As these VOCs are thought to reflect the physiological and metabolic status of the individual [5], they could represent tumor biology and response to therapy in individuals with cancer [5]. In addition, VOCs can be conveniently and reliably detected by Gas Chromatography and Mass Spectrometry (GC-MS), or gas sensors [21,44], which highlights the easy translatability of this application if successful. Taken together, the results of the previous studies coupled with the versality of VOC detection using GC-MS provided the framework to systematically study urinary VOCs as potentially biomarkers for PCa diagnosis. The model was developed and evaluated with the aid of statistical machine learning techniques and its performance was compared to the current standard, i.e., the PSA test.

2.2 Experimental section

2.2.1 Chemicals and Materials

All chemicals were of analytical grade. Mirex (99.0%, Dr. Ehrenstorfer GmbH, Germany), used as the internal standard, was purchased from the Laboratories of Dr. Ehrenstorfer, Germany.

Mirex solution of 100 mgL^{-1} was prepared in methanol (LCMS grade, Burdick & Jackson (Muskegon, MI, USA)). Hydrochloric acid (HCl, 37%) was purchased from Sigma–Aldrich (St. Louis, MO, USA). Ultra-pure deionized water from Milli-Q system (Millipore, Bedford, MA, USA) was used in the preparation of HCl solution and dilution of urine samples.

2.2.2 Patient recruitment and sample collection

Internal Review Board approvals and written informed consents were obtained for the multi-institutional study. Patients presenting to Geisinger Medical Center urologic clinic for evaluation of elevated PSA >2.5 ng/ml or abnormal digital rectal exam in the urology clinic were selected for the study. Patients who did not wish to participate in this study or whose urinalysis was suspicious for infection were excluded. All patients provided urine specimens for dipstick urinalysis to rule out infection prior to office-based transrectal ultrasound guided prostate biopsy by a single provider (HW). The remaining urine samples (50 ml) were collected and stored at -80^{o}C prior to VOC analysis. Patients with a positive PCa diagnosis represented the cancer cases. Those with benign prostate or low volume (≤2 of 13 cores involved) high grade prostate intraepithelial neoplasia or atypical small acinar proliferation (ASAP) were included in the control group so as to mimic the real world clinical scenario. The two patients with low volume ASAP had a negative repeat 6 month prostate biopsy. The study consisted of two cohorts: training and testing. In the training cohort, 108 men (aged from 40 to 84) were included (

Table 2. 1). Of the 108 men, 55 were diagnosed with PCa, while 53 were negative for PCa (herein, called controls). To validate the developed PCa diagnostic model, urine specimens from 53 PCa positive and 22 PCa negative patients (herein, called testing cohort) were obtained from an internist (MHA) with the Clinic Internal Medicine of El Paso, Texas. Specimens were obtained prior to patient referral to urology. Patients were subdivided into their respective groups based on prostate biopsy results as described for the training group

Table 2. 1 Demographic information of prostate cancer and cancer-negative patients in the VOC PCa diagnosis model development. Data are presented as median (interquartile range) for continuous variables and n (%) for categorical variables.

	Training Cohort (Model Development)			Testing Cohort (Model Evaluation)		
	Prostate cancer patients	Controls	p value [b]	Prostate cancer patients	Controls	p value [b]
N	**55**	**53**		**53**	**22**	
PSA[a] (ng/mL)	**5.29 (0.08-1987)**	**2.6 (0.1-18.2)**	**0.28**	**3.75 (0.08-75)**	**5.21 (1.8-20)**	**0.96**
Gleason score		**N/A**			**N/A**	
6	**20 (36%)**			**31 (58%)**		
7	**23 (42%)**			**16 (30%)**		
8	**6 (11%)**			**3 (6%)**		
9	**6 (11%)**			**3 (6%)**		

[a] PSA: prostate specific antigen;

[b] The p value from the t-test of the PSA numbers between prostate cancer and control groups

2.2.3 Extraction of VOCs from urine samples by Stir Bar Sorptive Extraction (SBSE)

Urine samples were thawed in ice, transferred to centrifuge tubes, and centrifuged for 10 minutes at 300g prior to extraction. To extract the organic metabolites (i.e. VOCs), 1.0 mL of urine sample supernatant, 19.0 mL of DI water, 300 µL of 100 ppm Mirex solution (internal standard) and 600 µL of 2 M hydrochloric acid were added into a 20 mL amber vial. A commercially available Stir Bar coated with polydimethylsiloxane (Twister™, 10 mm × 1 mm, Gerstel, Mülheim an der Ruhr, Germany) was then placed into the vial, and the solution was stirred for 2 hours at 1000 rpm. At the end of the stirring, the stir bar was removed from the solution, rinsed with DI water, dried with lint free paper, and placed into a thermal desorption tube for chemical analysis.

2.2.4 GC-MS analysis

Urinary VOCs were analyzed in a thermal desorption unit (TDU, Gerstel), coupled with a 6890 GC system and a 5973 N Mass Selective Detector (Agilent Technologies, Wilmington DE). The thermal desorption process was programmed as follows. The initial temperature was set at 45 °C holding for 0.5 min; the temperature was increased to 300 °C at 60 °C min^{-1} and held for 5 min. Desorption gas flow was set at 1.0 mL min^{-1}. During desorption, all the desorbed compounds were concentrated in a cold injection system, CIS4 (Gerstel), at -40 °C prior to GC injection. Once the desorption process was completed, the CIS4 was heated to 300 °C at 12 °C sec^{-1} and held for 5 min in a solvent vent mode. The VOCs were separated and analyzed by GC-MS using splitless mode through a ZB-5ms capillary column (30m × 0.25mm × 0.25µm, Phenomenex, USA). The oven temperature was programmed as follows: held at 35 °C for 5 min; heated to 300 °C at 10 °C min^{-1}, and held for 10 min. The VOCs were detected by Mass Selective detector in scan mode (20-500 m/z). The National Institute of Standards and Technology (NIST) Library was used for the identification of VOCs profile in urine sample.

2.2.5 Data processing and statistical analysis

We used mirex as the internal standard because of its non-existence in human urine. The relative intensity of each VOC peak could then be normalized against that of Mirex, allowing semi-quantitative statistical analysis of VOCs.

Over 9,000 different VOC types were found in the urine samples, resulting in a high-dimensional modeling problem. To streamline the analysis, we first removed the VOCs that could be observed in less than 3% of the entire population. The remaining variables were screened by testing the difference in each VOC between the PCa positive and control groups. The Wilcoxon rank-sum test was used since it can accommodate the zero inflation among many VOCs. Heat maps were generated to visualize those significant VOCs (p<0.05) in the PCa positive and control groups.

Applying a liberal cutoff of 0.2 to the p-values, over 800 VOCs remained for the model development. To deal with this p>>n scenario (i.e., numbers of VOCs are much greater than the number of samples), we fit regularized logistic regression models [136] with LASSO [137] penalty, and the 10-fold cross-validation was used to select the optimal tuning parameter. The final logistic model was then evaluated via the Receiver Operating Characteristic (ROC) curve and other performance measures on the basis of jackknife prediction [138], which helps alleviate the over-optimism induced by variable selection. Furthermore, Firth's approach was taken to fit the final logistic model in order to achieve bias-reduction for the small sample scenario and deal with the nearly complete separation seen in the data [139]. Another R package, known as OptimalCutpoints, was used to determine the optimal cut point for the diagnostic model corresponding to the maximum Youden Index [140]. All statistical analyses are performed using the open-source statistical computing software *R* [141].

2.3 Results

The study analyzed VOCs in urine samples collected from patients to develop the urine metabolome based PCa diagnosis model. All VOCs were identified using an existing NIST library

and significant VOCs were selected based on their occurrence and relative quantity in the urine. One mL of urine sample was extracted by stir bar sorptive extraction (SBSE) , and the VOCs were then analyzed by GC/MS. The quantity of each VOC was normalized to the internal standard, mirex, by taking the ratio between the signal of the compound and that of mirex.

A total of 9,144 potential VOCs were detected in urine from 108 patients (55 PCa positive and 53 PCa controls). Using the Wilcoxon test, 254 VOCs were found to be positively related to PCa and 282 VOCs were negatively associated at $p<0.05$ (Figure 2. 1 a).

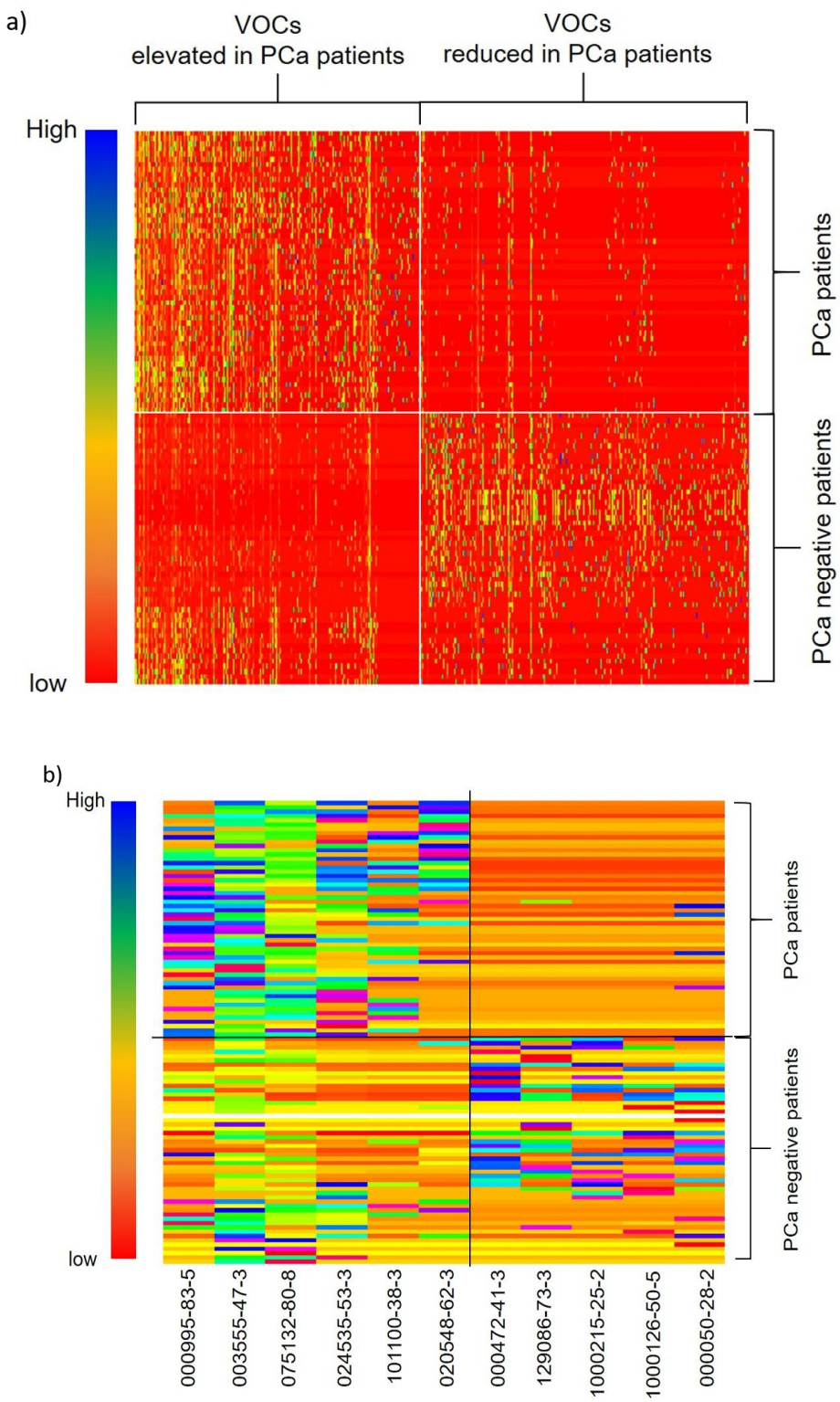


Figure 2. 1: a) The heat map of significant VOCs by Wilcoxon test ($p < 0.05$) in PCa vs PCa negative samples. The correlation between VOCs and patients ranges from low (red) to high (blue); b) the heat map of 11 selected VOCs in the urinary VOCs based PCa diagnostic model.

The values of those selected VOCs in patients show a strong pattern in distinguishing PCa positive and negative patients.

To avoid missing important predictors for PCa prevalence, a relatively large (with $p<0.20$) was applied to screen variables for further development of the regression resulting in a total of 850 potential VOCs. After l_1 regularization, the final logistic model contains 11 VOCs (listed in the

Table 2. 2), and a heat map was generated to show their differentiation in PCa positive and negative patients (Figure 2. 1 b). On the basis of predicted probabilities from the final model obtained via jackknife cross-validation, the area under the receiver operating characteristic (ROC) curve (AUC) is 0.92 with sensitivity of 0.96 and a specificity of 0.80 respectively. As a comparison, we also built a logistic model with PSA only, which yielded an AUC of 0.54. The model resulted in a sensitivity at 0.60 and a specificity at 0.42 respectively when applied to the test data (Figure 2. 2 b). Testing the performance of the PCa diagnostic model using an external cohort of 75 patients (53 PCa patients and 22 PCa negative patients) yielded a ROC curve of of 0.86 (Figure 2. 3), with an 87% sensitivity and a 77% specificity.

Table 2. 2. The VOCs selected by logistic regression models for prostate cancer diagnosis prediction and risk assessment.

CAS Number[1]	Formula	Chemical Name	Dominating Group	p value[4]	Occurrence	
					PCa(+)[2]	PCa(-)[3]
000472-41-3	$C_{18}H_{20}O_2$	4-(3,4-dihydro-2,2,4-trimethyl-2H-1-benzopyran-4-yl)-phenol	PCa negative	1.29E-07	0	22
000050-28-2	$C_{18}H_{24}O_2$	Estradiol	PCa negative	9.38E-05	4	20
129086-73-3	$C_{16}H_{32}O_3$	Ethyl à-hydroxymyristate trisiloxane	PCa negative	4.64E-06	1	19
1000215-25-2	$C_{15}H_{22}O$	1-(2,4-Dimethylphenyl)-3-(tetrahydrofuryl-2)propane	PCa negative	2.54E-05	0	15
1000126-50-5	$C_4H_5N_3O_2$	2-amino-Imidazole-5-carboxylic acid	PCa negative	2.54E-05	0	15
000995-83-5	$C_{10}H_{32}O_4Si_5$	1,1,3,3,5,5,7,7,9,9-decamethyl-pentasiloxane	PCa positive	1.21E-06	35	9
003555-47-3	$C_{12}H_{36}O_4Si_5$	1,1,1,5,5,5-hexamethyl-3,3-bis[(trimethylsilyl)oxy]-Trisiloxane	PCa positive	1.44E-06	52	37
020548-62-3	$C_{26}H_{42}O_4$	Phthalic acid, bis(7-methyloctyl) ester	PCa positive	0.000388	23	8
024535-53-3	$C_{12}H_8ClNO_3S$	4-Nitro-4'-chlorodiphenylsulfoxide	PCa positive	8.06E-06	46	23
075132-80-8	$C_3H_7Cl_5N_3P_3$	1-Propylpentachlorotriphosphazene	PCa positive	5.16E-06	47	36
101100-38-3	$C_{15}H_{22}O_2$	2,6-di-t-butyl-4-hydroxymethylene-2,3,5,6-detetrahydrocyclohexanone	PCa positive	4.32E-05	23	4

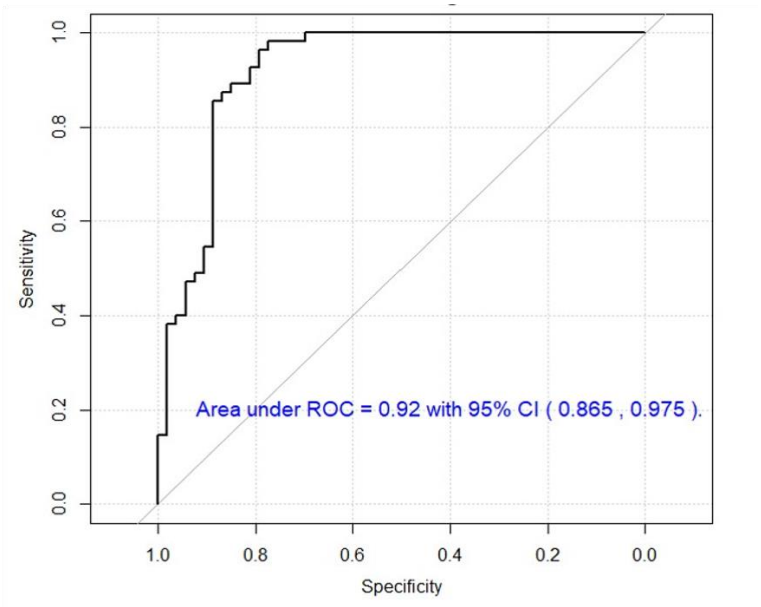
1: Chemical Abstracts Service number

2: PCa(+): prostate cancer positive patients

3: PCa(-): prostate cancer negative patent

4: p value: the p value of selected compounds from Wilcoxon rank-sum test

a)



b)

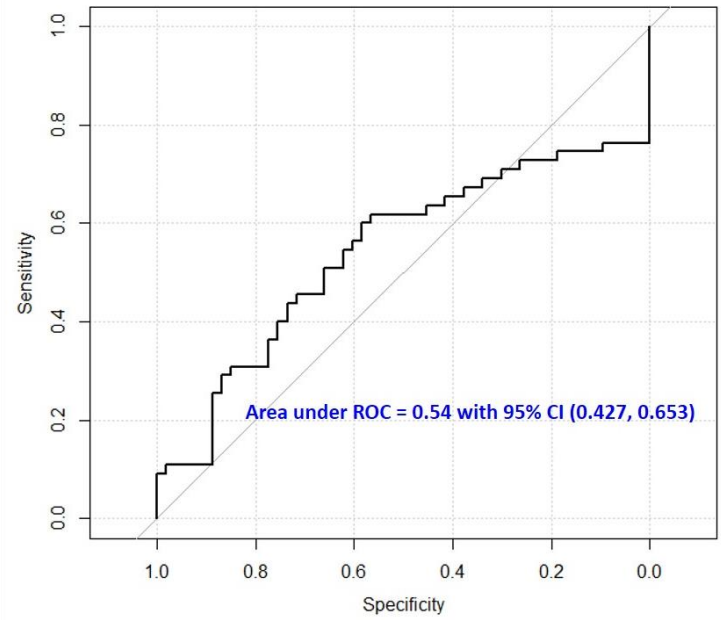


Figure 2. 2: a) The ROC curve for VOC PCa diagnosis logistic model verified in 108 patients; b) The ROC curve for logistic model with PSA only.

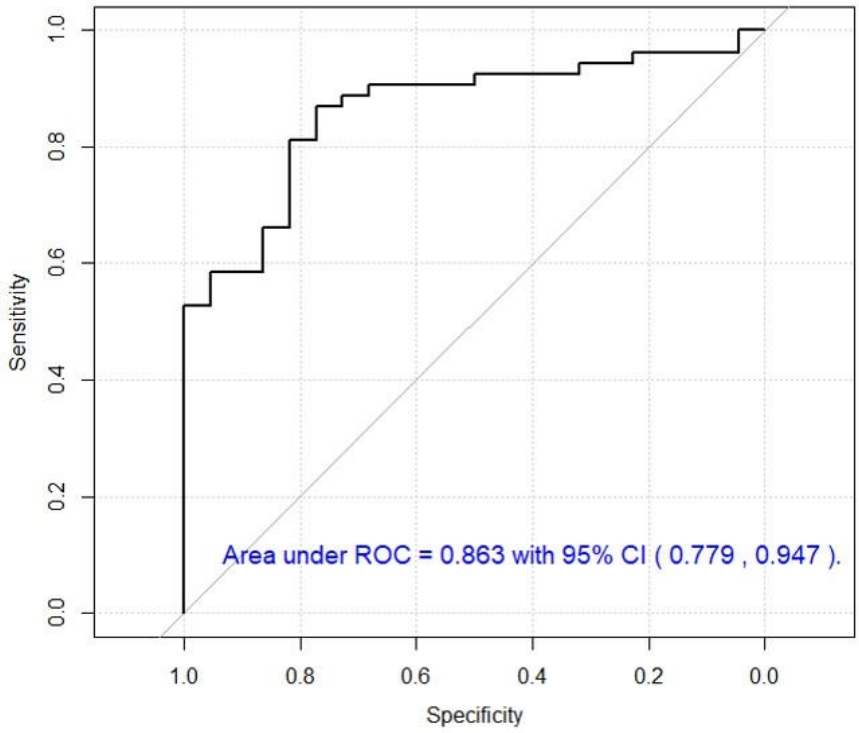


Figure 2. 3: The ROC curve for VOC PCa diagnosis logistic model tested in 75 patients

2.4 Discussion

There is significant interest in PCa biomarker development as evidenced by the number of biomarkers currently available for disease diagnosis, risk stratification, prognosis and response to treatment[142]. In this study, we developed a urinary VOCs based model for PCa diagnosis. The analytical method allowed for an easy, fast and efficient analysis of VOCs without tedious sample preparation. The solventless sample preparation technique preserved the sample integrity and permitted effective analyses for processing large number of samples, which is an important factor for clinical translatability. Unlike gas-sensors (e.g., E-nose) which are based on signal indications [44], the GC-MS method identified relevant compounds which can be further validated or interrogated in future metabolomics and physiological studies. Using the internal standard, Mirex, facilitated the semi-quantitative determination of VOCs. The expression of VOCs as continuous data rather than as binary provides a more disease-specific representation of metabolite concentration (i.e increase or decrease) *in vivo*, and improves the subsequent model development.

For the urinary VOC based PCa diagnostic model, 11 VOCs were selected into the final logistic regression. The model was validated and produced an AUC of 0.92 (Figure 2. 2 a), suggesting the potential of urinary VOCs for PCa diagnosis. While other studies have evaluated various urinary analytic techniques for PCa detection, the study by Khalid et al represents the best study to date, utilizing urinary VOCs for this purpose. Their study was comprised of 59 PCa patients and 43 controls. Following urinary VOC analysis of prostate cancer positive and negative specimens, they manually selected 4 VOCs **(**2,6-dimethyl-7-octen-2-ol, pentanal, 3-octanone, and 2-octanone) for their PCa detection model. The accuracy of this model was 63-65% but improved to 74% and 65% when combined with PSA using their random forest and linear discriminant analysis methods respectively. [21]. In their study, PSA alone was marginally better than the 4 VOC model which call into question the clinical utility of these set of biomarkers. By comparison, in our study, there was a significant difference between the performance of PSA (AUC of 0.54) and the diagnostic model (AUC of 0.92). Manually selection of VOCs likely introduced bias into the model as opposed to allowing the discriminatory analysis to dictate the final VOCs that would come to define each cohort. In our study, there was a blend of VOCs that were increased or reduced in either PCa positive or PCa negative groups. In the Khalid study, the same cohort was used for both discovery and validation. Despite using robust statistical methods to minimize the bias associated with this approach, the sensitivity and specificity of the 4 VOC remained poor. In this study, in spite of using a rural Pennsylvania patient cohort to develop the PCa detection model, the validity of the model was upheld when tested in an urban Texas population with an AUC of 0.86 with sensitivity and specificity of 0.87 and 0.77 respectively.

Table 2. 3. Comparison in sensitivity, specificity, and AUC from various biomarkers in prostate cancer diagnosis

	Our study			Potential biomarkers					
	VOCs based model		PSA	Iso-PSA	PCA3	TMPRSS2 :ERG	4Kscore	PHI	ConfirmMDx
	Validation	Evaluation							
Sensitivity	**0.96**	**0.87**	**0.60**	**0.90**	**0.68**	**0.24**			**0.68**
specificity	**0.80**	**0.77**	**0.42**	**0.48**	**0.58**	**0.93**			**0.64**
AUC	**0.92**	**0.86**	**0.54**	**0.79**	**0.68**	**0.59**	**0.82**	**0.68**	

Several other biomarkers, including Iso-PSA, prostate cancer antigen 3 (PCA3), 4Kscore, Prostate Health Index (PHI), TMPRSS2:ERG and ConfirmMDx [66-71], were developed as alternatives to address the challenges of PSA test in PCa diagnosis. Among those markers, IsoPSA, PHI and 4Kscore are PSA-based assay for PCa risk assessment [66,69,71]. PCA3 is reported to be a noncoding RNA that is prostate specific and highly overexpressed in PCa [72]; and TMPRSS2-ERG gene fusion is the predominant molecular subtype of PCa [73]. ConfimMDx is an epigenetic test for PCa diagnosis before prostate biopsy [70,143]. However, these PCa diagnostic tests are still limited by either their low sensitivity and/or low specificity (

Table 2. 3). The metabolome can be considered as the amplified output of the biological system, and therefore metabolomics profiling is emerging as a promising strategy for PCa diagnosis [144]. In this study, the VOCs metabolome based PCa diagnosis model outperformed those aforementioned diagnostic PCa biomarkers (

Table 2. 3).

Among the 11 compounds selected in the PCa diagnosis model, 5 showed higher levels in the PCa negative group), while the remaining 6 were positively related with PCa. Some of these compounds may be directly or indirectly involved in androgen receptor (AR) signaling. For example, androgens and AR in prostate regulate the activity/expression of the enzymes involved in lipogenesis [96]. Lipogenesis is mediated by increased activities of fatty acid biosynthetic enzymes (including ACC1, FASN, and stearoyl CoA desaturase (SCD1)) [114-116]. The urinary fatty acids and/or fatty acid metabolites found in this study could be a reflection of such activities and therefore relevant in PCa.

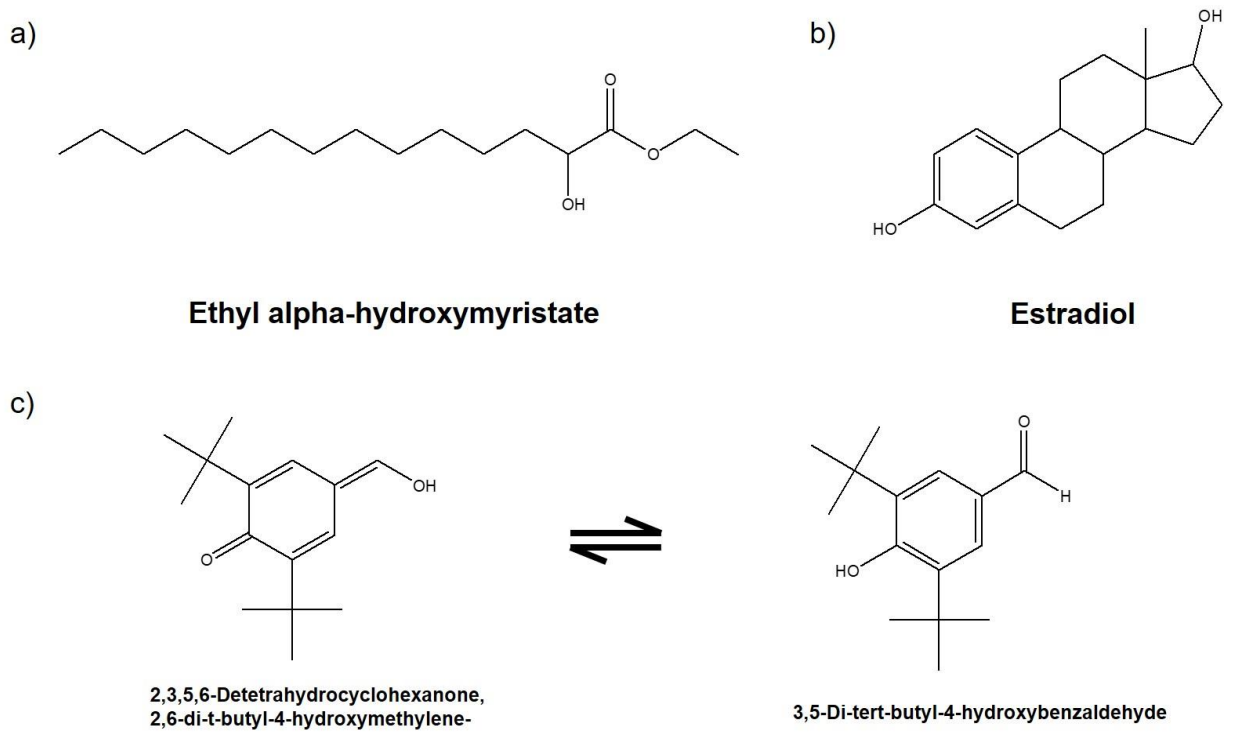


Figure 2. 4. Chemical structures of four representative VOCs selected by the regression model

Ethyl α –hydroxymyristate (Figure 2. 4), was reduced in PCa patients. Numerous epidemiological studies suggest that myristic acid (MA) significantly increase the risk of prostate cancer–specific mortality among patients diagnosed with localized disease [145,146]. *In vivo* studies

in mice showed that myristoylation of Src kinase promoted Src-induced and high-fat diet–accelerated prostate tumor progression[147]. The low levels of ethyl α-hydroxymyristate seen in PCa patients urine samples from could be linked to the consumption of this metabolite during the PCa tumorigenesis process. However, this would need to be validated experimentally. Moreover, ethyl α –hydroxymyristate, contains the moiety of 2-hydroxymyristic acid which was reported to inhibit the myristoylation and alters the stability of $p56^{lck}$in T cells. $p56^{lck}$ is a protein-tyrosine kinase that is found predominantly in lymphoid cells [148], and was reported to be positively expressed in PCa cell lines and tissues [149], and in metastatic cancer [150].

We also found 2,6-di-*t*-butyl-4-hydroxymethylene-2,3,5,6-detetrahydrocyclohexanone to be elevated in the PCa. This compound is the tautomer of 3,5-di-*t*-butyl-4-hydroxybenzaldehyde (Figure 4c), a component of the qHTS assay used to identify small molecule agonists of the RXR (retinoid X nuclear receptor alpha) signaling pathway: Summary (AID 1159531) [151]. The nuclear expression of RXR alpha receptor subtype was reported to be generally downregulated in human PCa cell lines and specimens, and the loss or reduction of RXR alpha function is a critical determinant in prostate tumorigenesis [152]. The high level of 2,6-di-*t*-butyl-4-hydroxymethylene-2,3,5,6-detetra-hydrocyclohexanone in PCa positive urine samples may reveal the low utilization of all those types of agonist of RXR signaling pathway due to the loss of RXR alpha function. Finally, the low level of estradiol in PCa positive urine samples is consistent with previous studies identifying low circulation levels of estradiol as a PCa risk factor [153].

There are several limitations associated with this study. Due to the random nature of prostate biopsies, it is plausible that PCa negative patients could be misclassified if they harbored disease missed on prostate biopsy. Secondly, the heterogeneous nature of PCa suggests that the urinary VOC PCa diagnostic model identified in this study will likely improve as sample size is increased and more patients from diverse backgrounds and environmental exposures are further assessed. While this study focused on urinary VOCs, the role of non-VOC products of metabolism may also represent important biomarkers that should be accounted for as the final comprehensive PCa diagnostic model is developed. Finally, the impact of time of urine

collection, diet, disease risk, personal genetic profile and environmental exposures on VOC concentration measured is unknown but will be addressed in future studies.

2.5 Summary

In summary, the urinary VOCs based model shows promise as a non-invasive and reliable method for PCa diagnosis. The evaluation in an external cohort of patients, demonstrated the high discriminating power of this strategy in PCa diagnosis. The urinary VOC based metabolites approach could likely be adapted into a clinically viable, highly sensitive, cost-effective portable diagnostic assay for PCa. The validity and effectiveness of those selected VOCs in PCa diagnosis need to be further confirmed in a larger cohort study. Though the investigation of the originating pathways of the VOCs selected in the model is beyond the scope of this study, some VOCs linked to AR modulated pathways were identified in our studies, suggesting a molecular rationale for the production of these VOCs related to PCa. The biological and chemical significance of these VOCs could be further studied to provide supporting information for the upstream proteomic pathways.

Chapter 3: Application of Volatile Organic Compounds (VOCs) in Prostate Cancer Risk Assessment

- Prostate cancer is a heterogeneous disease, ranging from indolent to life threatening stages.
- The study was to evaluate the performance of VOCs based model in risk assessment of prostate cancer. A PCa risk stratification model was developed and validated using innovative statistical machine-learning techniques.
- A VOCs profile screening showed the potential of VOCs for PCa risk stratification.

Abstract:

Background: Prostate cancer (PCa) is a heterogeneous disease ranging from indolent to life threatening stages. The choice of watchful waiting or therapies is always depending on the risk level of PCa. The performance of VOCs has been proved in our previous study to be able to strongly discriminate PCa patients from controls. We hypothesized that urinary volatile organic compounds (VOCs) could be used to develop urinary VOC PCa prognostic models for cancer risk stratification.

Methods: Urine samples from 89 men who presented for transrectal ultrasound guided prostate biopsy for an elevated PSA or abnormal digital rectal exam were collected. Based on the D'Amico risk scores system, these PCa patients were divided into two groups: low-risk group (indolent PCa, GS = 6, PSA< 10, n=55), and high/intermediate-risk group (clinically significant PCa, GS = 6 and PSA $\geqslant$ 10, or GS > 6 with any PSA values, n=34). Urinary metabolites were analyzed by Gas Chromatography-Mass Spectrometry. Urinary VOCs were screened for significance. A PCa risk stratification model was developed and validated using innovative statistical machine-learning techniques.

Results: Using the Wilcoxon rank sum test, 23 VOCs were found to be positively related to high/intermediate−risk PCa and 44 VOCs negatively associated ($p<0.05$). Regularized logistic regression analysis and the Firth method were then applied to predict PCa risk level, resulting in a final model that contains 11 VOCs. Under cross-validation, the area under the receiver operating characteristic curve (AUC) for the final model was 0.86 (sensitivity: 0.85; specificity: 0.80), which indicates a highly promising discrimination power of urinary VOCs in PCa high/intermediate risk assessment.

Conclusion: The study describes the development of a urinary VOC-based model for PCa risk stratification and the model shows the potential of VOCs for PCa prognosis.

Clinical Practice Points

- Due to the low specificity of current prostate cancer screening with serum Prostate-Specific Antigen (PSA), up to 80% of men who are found to have elevated PSA value and subsequently undergo prostate biopsies will either not be diagnosed with prostate cancer (PCa) or be diagnosed with an indolent form of PCa, which is unlikely to cause significant morbidity or mortality. It is urgent to address the problem of over-diagnosis and overtreatment in prostate cancer care.
- We developed urinary metabolites (i.e. VOCs) based model to assess the risk level for PCa with AUC 0.86, which indicated the great discrimination power of VOCs base model for PCa risk stratification.

3.1 Introduction

In United States, prostate cancer (PCa) is the most common non-cutaneous cancer and the second leading cause of cancer death in men.[59] In most cases, men, who are diagnosed with PCa, will ultimately die of other causes. And even with eventually lethal cases, the natural history of the disease is relatively protracted.[154] PCa is a heterogeneous disease ranging from indolent to life threatening stages. The diagnosis and treatment of PCa is based on a series of clinical states (Figure 3. 1), which begin with localized tumors, followed by the non-castrate state (including rising PSA state and metastatic state), and finally move to the castration-resistant states.[155,156] According to a decision analysis of alternative treatment strategies for clinically localized PCa carried out by Fleming et al.[157], the choice of watchful waiting is a reasonable alternative to the invasive treatment for men with localized PCa tumors. However, men with high-risk PCa, who account for 15% of all PCa diagnosis[158], are definitely in need of therapies.

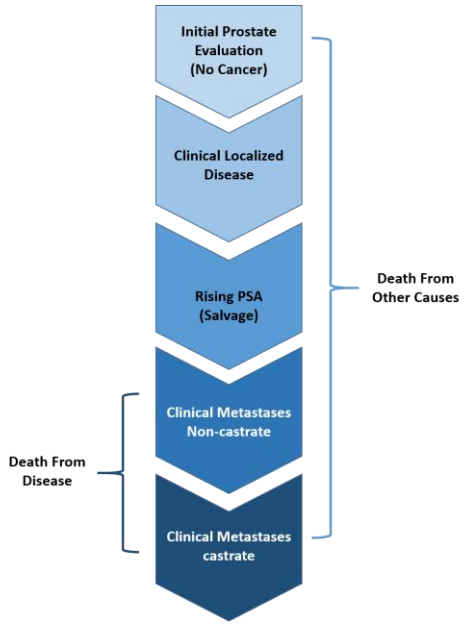


Figure 3. 1 Clinical states of prostate cancer progression.

The clinicians and researchers have increasingly recognize the importance of risk stratification of PCa in terms of indolent prostate carcinoma behavior and potential risks of available treatment[159], and the risk-adapted treatment strategies could then be implemented accordingly. Different risk factors associated with PCa have been integrated into individualized risk prediction, including PSA, Gleason score, T stage, and other risk factors[160]. In 1998, D'Amico et al. developed a combined modality staging system, which stratified PCa patients into groups with a low, intermediate, or high-risk of biochemical recurrence after radical prostatectomy or radiotherapy, according to the clinical stage, biopsy Gleason score, and preoperative prostate specific antigen level. However, the aforementioned risk-assessment factors require invasive procedure and the accuracy of PSA is questionable. Therefore, there is a significant interest in finding more sensitive, reliable and cost effective PCa screening and prognosis biomarkers.

VOCs are generated from the human body and thought to reflect the physiological and metabolic status of the individual [5,135]. Thus, VOCs in individuals with cancer could represent tumor biology and response to therapy [5]. In a recent study with trained dogs, the animals could smell the urine of men, and discriminate between men with and without PCa with 91% sensitivity

and specificity [8]. The odor is represented by a group of volatile components, i.e. VOCs, and VOCs profiles can be easily detected using gas chromatography-mass spectrometry (GC-MS), or gas sensors. [21,44] In our previous study, the performance of VOCs was proven to be able to strongly discriminate PCa patients from controls.[38] It raised the next question whether VOCs can be used in risk levels differentiation. Herein, we present the first study of the performance of VOCs in PCa risk stratification.

3.2 Experimental section

3.2.1 Chemicals and Materials

All chemicals, including internal standard (mirex), HCl, methanol, deionized water, were of analytical grade.[38]

3.2.2 Patient recruitment and sample collection

In this study, a total of 89 men who presented for transrectal ultrasound guided prostate biopsy for an elevated PSA above 2.5 ng/ml or abnormal digital rectal exam were included. Based on the Gleason score (GS) and PSA, these PCa patients were divided into two groups: low-risk group (GS = 6, PSA< 10) and high/intermediate-risk group (GS = 6 and PSA ≥ 10, or GS > 6 with any PSA values) as shown in Table 3. 1. The high/intermediate-risk group was considered to be clinically significant and low risk group as indolent PCa. Urine samples were collected at the medical facilities and stored at -80 ºC until chemical analysis.

Table 3. 1 Demographic information of prostate cancer patients in the VOC PCa risk assessment model study. Data are presented as median (interquartile range) for continuous variables and n (%) for categorical variables.

	Prostate cancer patients		p value[a]
	Low risk	High and intermediate risk	
N	34	55	
PSA (ng/mL)	5.29 (0.08-1987)		
	3.95 (0.10–9.33)	6.21 (0.08-1987)	0.22
Gleason score	6-9		
6	34 (100)		
7		38 (69)	
8		11 (20)	
9		6 (11)	

a. The p value from the t-test of the PSA numbers between low grade and high risk group.

3.2.3 Extraction of VOCs from urine samples

Urine samples were processed through centrifuging and stir bar sorptive extraction.[38]

3.2.4 Gas Chromatography-Mass Spectrometry analysis

Briefly VOCs from urine samples were analyzed in a thermal desorption unit, TDU (Gerstel), coupled with a Gas Chromatography-Mass Spectrometry (GC/MS) system. The National Institute of Standards and Technology (NIST) Library was used for the identification of VOCs profile in urine sample.[38]

3.2.5 Data processing and statistical analysis

We have implemented internal standard method in the analysis. Mirex was used as the internal standard for its non-existence in urine. The relative intensity of each VOC peak could then be normalized against that of Mirex to enable semi-quantitative analysis of VOCs in the statistical analysis. The statistical significance of each VOC was tested by Wilcoxon test. Heat maps were

generated to visualize significant VOCs ($p<0.05$) among the high/intermediate risk group and low risk group. Applying a liberal cutoff at $p=0.2$, a larger size of VOCs were included to develop a logistic regression model for PCa risk assessment. The VOC based model was then evaluated by cross-validation.

Briefly, the VOC-based diagnostic tool was developed via logistic regression. The final logistic model was evaluated via the Receiver Operating Characteristic (ROC) curve and its associated performance was measured on the basis of its jackknife prediction. All the analyses are performed using the open-source statistical computing package R.[38]

3.3 Results

A total of 89 PCa patients (55 High/intermediate risk PCa patients and 34 low risk PCa patients) were recruited for the study. There were 8,194 VOCs found in urine from the subjects. Using the Wilcoxon rank sum test, 23 VOCs were found to be positively related to high−risk PCa and 44 VOCs negatively associated at $p<0.05$, as shown in Figure 3. 2. After variable screening, 289 potential VOCs ($p < 0.2$) were selected for model development. The l_1 –regularized logistic regression resulted in a final model that included 11 VOCs as listed in the Table 3. 2. On the basis of predicted probabilities computed via Jackknife, the area under the receiver operating characteristic (ROC) curve is 0.86 as shown Figure 3. 3, and the sensitivity and specificity were 85% and 80% respectively, which indicates a highly promising discrimination power of urinary VOCs in PCa high risk assessment.

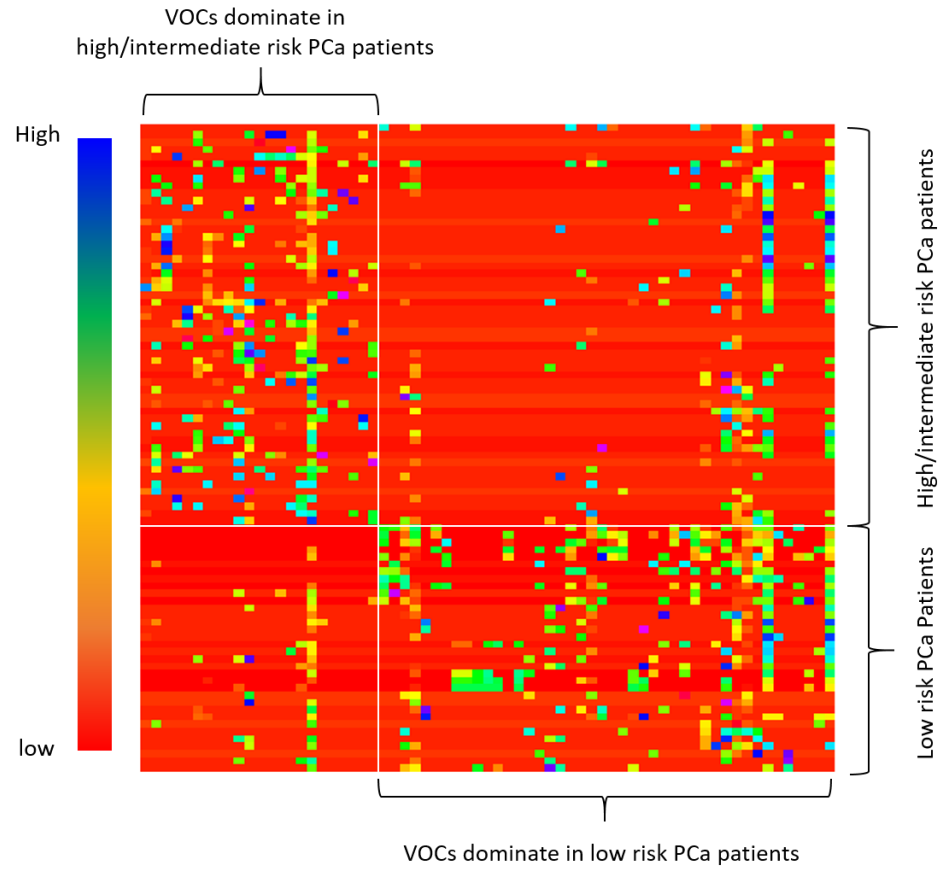


Figure 3. 2 The heat map of significant VOCs by Wilcoxon test ($p < 0.05$) in high/intermediate risk vs low risk prostate cancer groups. The correlation between VOCs and patients ranges from low (red) to high (blue).

Table 3. 2 VOCs from logistic regression models for prostate cancer risk assessment

CAS Number	Formula	Chemical Name	Dominating Group	Occurrence	
				High/intermediate risk	Low risk
031061-61-7	$C_{13}H_{20}O_2$	Tricyclo[4.3.1.1(3,8)]undecane-3-carboxylic acid, methyl ester	Low Risk	2	5
018127-01-0	$C_{13}H_{18}O$	4-(1,1-dimethylethyl)-benzenepropanal	Low Risk	1	5
007206-21-5	$C_{18}H_{36}$	5-Octadecene, (E)-	High Risk	14	15
020607-72-1	$C_6H_{14}N_2$	Acetaldehyde, butylhydrazone	Low Risk	19	3
046498-17-3	$C_{13}H_{13}N_3$	3,6-Diamino-9-methylcarbazole	High Risk	0	5
000111-06-8	$C_{20}H_{40}O_2$	Hexadecanoic acid, butyl ester	Low Risk	3	8
131316-14-8	$C_{17}H_{16}OS$	trans-3'-Methyl-4-(methylthio)chalcone	Low Risk	1	9
092617-73-7	$C_{13}H_{18}O$	2-(1,1-dimethyl-2-propenyl)-3,6-dimethyl-phenol	Low Risk	16	3
054340-85-1	$C_{12}H_{16}$	1-(2-butenyl)-2,3-dimethyl-benzene	Low Risk	2	6
001153-51-1	$C_{19}H_{30}O$	(3alpha,5alpha)-androst-16-en-3-ol	Low Risk	2	6
000621-42-1	$C_8H_9NO_2$	Metacetamol	Low Risk	1	5

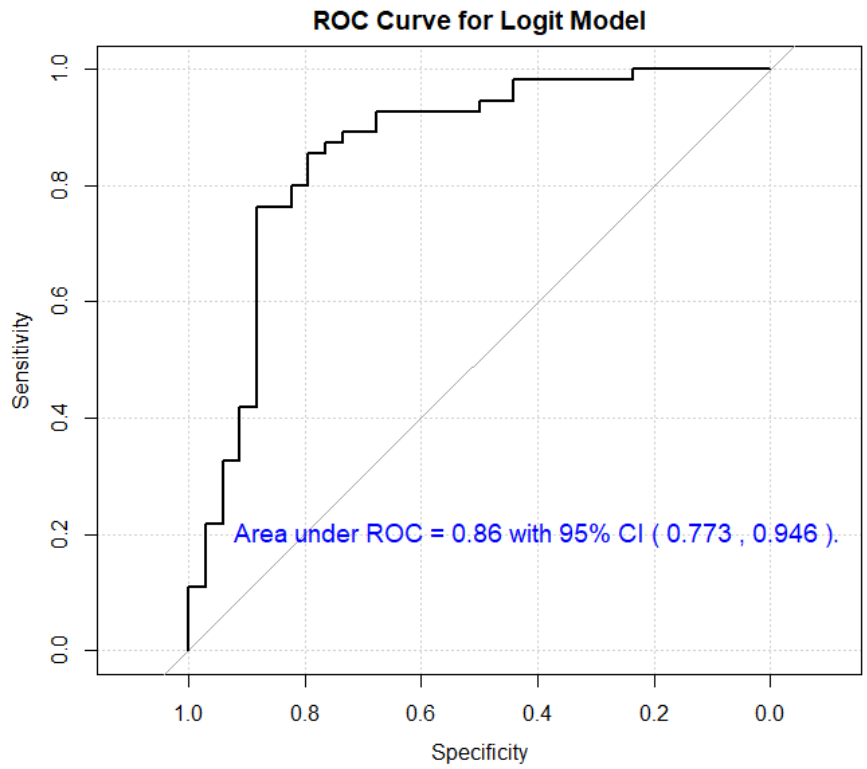


Figure 3. 3 The ROC curve for PCa risk assessment logistic model verified in 89 patients.

3.4 Discussion

In this study, the VOCs based PCa risk assessment model has exhibited the potential to discriminate PCa high/intermediate risk group patients from low risk PCa patients. Additional to the well-known risk factors, age, PSA, family history, DRE findings, and outcomes of prior biopsy, VOCs would be considered as another risk factor in PCa risk stratification.

The study of VOCs in PCa risk assessment has identified 11 VOCs in the final model with AUC 0.86, which has suggested a promising role of VOCs in PCa risk stratification. There are also other risk prediction tools being developed to implement the PCa diagnose and risk evaluation. Tomlins et al.[161] reported urine TMPRSS2:ERG fusion transcript, in combination with urine PCA3, could be used in PCa risk stratification of men with elevated serum PSA, with AUC ranging from 0.71 to 0.79. In the study Rastinehad et al.[162], using multi-parametric magnetic resonance

prostate imaging (MRI) may assess the degree of risk associated with magnetic resonance visible lesion in the prostate based on the D'Amico risk scores.

Among 11 VOCs selected by the regression mode of PCa high risk vs PCa low risk, 9 VOCs were negatively associated with escalating PCa risk and the remaining 2 were positively associated as listed in Table 3. 2. Androstenol, dominating in PCa low risk group, has a structure similar to that of androgens. Androstenol could be efficiently transformed from the precursor 5,16-androstadien-3β–ol (a sex steroid) through enzymatic activities including 3β-hydroxysteroid de-hydrogenase, 5α-reductase and 3α-hydroxysteroid dehydrogenase [163]. Of those three enzymes, 3β-hydroxysteroid de-hydrogenase and 5α-reductase were reported to be involved in PCa development and progression [164,165]. And studies also showed that androstenol could modulate the activity of CAR (constitutive androstane receptor) and PXR (pregnane X receptor, also known as steroid and xenobiotic receptor SXR) and the expression of some cytochrome P450 drug-metabolizing enzymes [166,167]. A combined analysis with CYP3A4 expression revealed that the low expression of the SXR and CYP3A4 was a significantly unfavorable prognostic factor for PCa in a multivariate analysis. It was suggested that the downregulation of the SXR, and consequently, its target CYP3A4 gene might play a significant role in PCa progression. Androstenol might exert tumor-inhibitory effects on PCa by increasing SXR expression and enhancing androgen clearance [168]. Therefore, the lower level of androstenol in the high risk group was consistent with the reported biological effect, i.e. relating the down regulation of SXR activities with the progression of PCa.

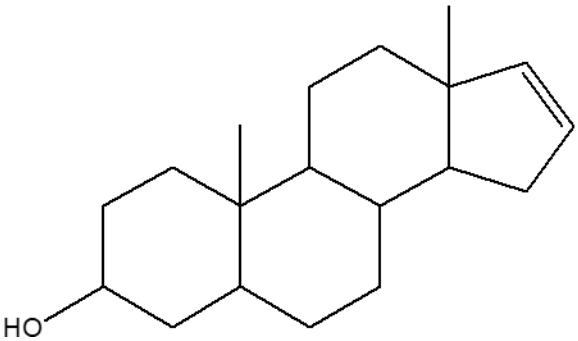


Figure 3. 4 The chemical structure of Androstenol**.**

3.5 Summary

Using VOCs in urine is a non-invasive and promising method for PCa risk assessment. The biological and chemical significance of some selected VOCs in this study could link to specific metabolic pathways in PCa progression and could be further studied to provide supporting evidence for the upstream proteomic pathways. Even with a considerably small sample size, the VOC models showed very promising discriminating power between high risk and low risk PCa. Our results showed that urinary VOCs could likely be translated into a clinically viable, highly sensitive, cost-effective portable diagnostic assay for PCa and may help identify patients with clinically significant PCa. For future study, the validity and effectiveness of those selected VOCs in PCa risk assessment may be further confirmed in a larger cohort.

Risk assessment is a critical aspect of PCa treatment decision-making. Hernandez et al. did a contemporary evaluation of the D'Amico risk classification of PCa.[169] They concluded from their study that the clinical relevance of this classification scheme may be limited and diminishing in the contemporary era. Besides the D'Amico classification, there are some other approaches to risk assessment of PCa, including a variety of nomograms[170] and UCSF-CAPRA Score[171]. The different risk classification methods would be applied in the further development and evaluation of VOCs based PCa model.

Chapter 4: Application of Urinary Volatile Organic Compounds (VOCs) for the diagnosis of Renal Cancer

- This chapter described the development of a diagnostic tool for clear cell renal cell carcinoma.
- Urine samples were collected from 108 patients who were imaging and/or surgery proven of clear cell renal cell carcinoma positive or negative.
- Urinary metabolites were analyzed to develop a metabolomics based model to detect clear cell renal cell carcinoma.
- The model was validated in a second cohort of patients and showed high discriminating power in clear cell renal cell carcinoma detection.

Abstract:

Background: Kidney cancer (renal cancer) accounts for more than 2% of cancer incidence and mortality in the United States The outcome of renal cell carcinoma (RCC) is usually unpredictable even after a long period of asymptomatically development and progression., its diagnosis is often incidental through the use of medical imagology and it is frequently at an advanced stage and metastatic when detected. Therefore, a fast and reliable early screen of renal cell carcinoma (RCC) enables better treatment outcome and care in patients. Clear cell RCC accounts for 70–80% of all kidney cancers. However, there is no recommended screening tests for RCC available clinically. This study developed a non-invasive and fast urine based diagnostic tool and evaluated its clinical utility, sensitivity and specificity to screen clear cell RCC (ccRCC, the main type of RCC) in populations.

Methods: Two cohorts were recruited for the study – a training cohort for model development and a testing cohort for model validation. For the training cohort, a total of 108 urine samples were obtained from 71 patients who were undergoing partial or radical nephrectomy and 37 patients ccRCC negative based the imaged renal mass. Specific volatile organic compounds (VOCs) in urine were analyzed by Stir Bar Sorptive Extraction coupled with Thermal Desorption-Gas Chromatography/Mass Spectrometry. All VOCs were analyzed based on their occurrence and relative quantity in the urine. The significant VOCs were screened by Wilcoxon Rank Sum Test. A VOCs based ccRCC diagnostic model was developed through the logistic regression in training group (57 ccRCC vs 31 controls) and validated in the testing group (14 ccRCC vs 6 controls).

Results: A total of 8,266 VOCs were found in the urine samples of training set. Using Wilcoxon Rank Sum Test in screeing their bivariate association with ccRCC, 79 VOCs were found to be related to urine samples of ccRCC positive patients while 91 VOCs corresponding to RCC negative controls ($p<0.05$). After further selection with l_1 regularization, 15 VOCs were included in the RCC diagnostic logistic model. On the basis of predicted probabilities from the model via cross-validation, the area under the receiver operating characteristic curve (AUC) was found to be 0.87 and the sensitivity and specificity were 93% and 77% respectively. The VOCs

based RCC diagnostic model were then validated in the testing group. The results showed a promising diagnostic power of this VOCs model in ccRCC screening with 0.81 AUC, 86% sensitivity and 83% specificity respectively.

Conclusion: This is the first systematic study to demonstrate and validate the clinical utility of a non-invasive urinary VOCs based diagnostic model in ccRCC screening. The VOCs based diagnostic model has the substantial potency and clinical value in RCC screening, and the analytical method was fast and highly translatable.

Clinical Practice Points

- No early screening method is recommended to screen for kidney cancer in people at average risk or increasing risk.
- We developed urinary metabolites (i.e. VOCs) based model to detect ccRCC with a high accuracy (87.5%) in ccRCC screening, thus potentially providing a new tool to accurately and reliably diagnose clear cell renal carcinoma.
- As a non-invasive and sensitive tool, the VOC model is ideally suited for early diagnosis of ccRCC. Future validation of our model will enable clinically relevant and cost-effective, point-of-care ccRCC screening and prognosis with the potential to save health care costs.

4.1 Introduction

Kidney cancer accounts for more than 2% of cancer incidence and mortality in the United States, which would include nearly 65,340 new cases (3.8%) and 14,970 deaths estimated for 2018.[59] The most common type of kidney cancer is renal cell carcinoma (RCC), about 9 RCC out of 10 kidney cancers. RCC is a heterogeneous malignancy, both morphologically and genetically, which is classified into different histologic subtypes, including clear cell RCC (ccRCC), papillary RCC, chromophobe RCC and other less common subtypes.[74-76] The most common subtype is ccRCC, accounting for 70–80% of all kidney cancers, which is so named because the high lipid content in the cytoplasm is dissolved during histological preparation methods leaving a clear cytoplasm.[172]

The outcome of RCC is usually unpredictable even after a long period of asymptomatically development and progression.[77] Therefore, its diagnosis is often incidental through the use of medical imagology and it is frequently at an advanced stage and metastatic when detected clinically.[78] However, no early screening method is recommended to screen for kidney cancer in people at average risk or increasing risk.

With the recognition of reprogramming metabolism as a central hallmark of cancer, metabolomic approaches have shown promise in oncology diagnosis and treatment.[129] Therefore, the applications of metabolomics are increasingly in the discovery and validation of biomarkers for cancers, which might not only improve early detection but also enable the outcome and recurrence prediction of disease. In a recent studies, trained dogs could differentiate men with and without cancers, like bladder cancer, prostate cancer and lung cancer, by sniffing their urine with high sensitivity and specificity.[8-10] The odor of urine is from so-called volatile organic compounds (VOCs) which can be generated from the human body and released through breath, blood, skin, urine and feces.[135] These VOCs are thought to reflect the physiological and metabolic status of the individual. [5] Thus in individuals with cancer, the VOCs could represent tumor biology and response to therapy.[5] In general, RCCs are epithelial tumors in contact with the urinary space[173], making this cancer well suited for a metabolomic approach through the VOCs analysis of urine.

VOCs can be easily detected using gas chromatography-mass spectrometry (GC-MS), or gas sensors.[21,44] We had demonstrated a successful application of VOCs in prostate cancer screening and prognosis.[38] The purpose of this study was to test the hypothesis that urinary VOCs based diagnostic model could be used in ccRCC early screening.

4.2 Experimental section

4.2.1 Study Design

Approval was obtained from the Internal Review Board (UTEP and Geisinger), and written informed consent was obtained from all patients. For ccRCC diagnostic model development, 108 urine samples in total were obtained from a) 71 patients preoperatively on the day of surgery who were undergoing partial or radical nephrectomy with a presumptive diagnosis of ccRCC based on a CT imaged renal mass (and whose postoperative pathology diagnosis established clear cell renal carcinoma cancer); and b) 37 patients RCC negative based the imaged renal mass.

Patients were divided into two groups: training group (for model development) and testing group (for model evaluation). The training set contained urine samples from 57 pathologically diagnosed ccRCC patients and 31 RCC negative control patients. For the testing group which was to evaluate the final performance of the ccRCC diagnostic model, 14 ccRCC patients and 6 RCC negative controls were involved. The demographic data are shown in Table 4. 1. Urine samples of the subjects were collected at the medical facilities and stored at -80 ^{o}C until chemical analysis.

Table 4. 1. Demographic information of clear cell renal cell carcinoma and renal cancer-negative patients in the VOC ccRCC diagnosis model development. Data are presented as median (interquartile range) for continuous variables and n (%) for categorical variables.

	Training Cohort (Model Development)		**Testing Cohort (Model Evaluation)**	
	Cancer Group	**Control Group**	**Cancer Group**	**Control Group**
No.	57	31	14	6
Age	64 (40-90)	64 (40-90)	63.5 (52-82)	53 (38-65)
Gender				
M	36	23	10	6
F	21	8	4	0
Tumor grade*		N/A		N/A
1	3 (5%)		0 (0%)	
2	24 (42%)		8 (57%)	
3	18 (32%)		4 (29%)	
4	12 (21%)		2 (14%)	

* **Pathology/Biopsy Confirmed Dx (Tumor grade)**

4.2.2 Chemicals and Materials

Briefly, all chemicals, including mirex, HCl, methanol, deionized water, used were of analytical grade.[38]

4.2.3 Extraction and Chemical analysis of VOCs from urine samples

Urine samples were processed through centrifuging and stir bar sorptive extraction.[38] Briefly, VOCs from urine samples were analyzed in a thermal desorption unit, TDU (Gerstel), coupled with a GC/MS system. The National Institute of Standards and Technology (NIST) Library was used for the identification of VOCs profile in urine sample. All samples were analyzed in a blinded and coded fashion in instrument analysis.[38]

4.2.4 Data processing and statistical analysis

The detailed analysis was previously published.[38] Mirex was used as the internal standard. The relative intensity of each VOC peak could then be normalized against that of Mirex to enable semi-quantitative analysis of VOCs in the statistical analysis. The statistical significance of each VOC was tested by Wilcoxon test. Heat maps were generated to visualize significant VOCs ($p<0.05$) among the ccRCC positive and control groups. Applying a liberal cutoff at $p=0.2$, a larger size of VOCs was applied to develop a logistic regression model was applied for further selection of noteworthy VOCs.

The VOC-based diagnostic tool was developed via logistic regression [136]. The final logistic model was evaluated via the Receiver Operating Characteristic (ROC) curve and its associated performance was measured on the basis of its jackknife prediction (Kleinbaum and Klein, 2010 [138]). All the analyses are performed using the open-source statistical computing package R.[141]

4.3 Results

All VOCs were identified based on their occurrence and relative quantity in the urine. The relative quantity of each VOC was determined through being normalized by the spiked Mirex (the internal standard, IS, for it has no inference with the VOCs *in vivo*.)

A total of 8,266 potential VOCs were detected in urine collected from the training cohort which consisted of 57 ccRCC patients and 31 age matched negative controls. Using the Wilcoxon rank sum test at statistical significance $p=0.05$, 79 VOCs were found to be related to cancer group

urine samples and 91 VOCs corresponding to controls. The distribution of those selected VOCs in patients was shown in Figure 4. 1a).

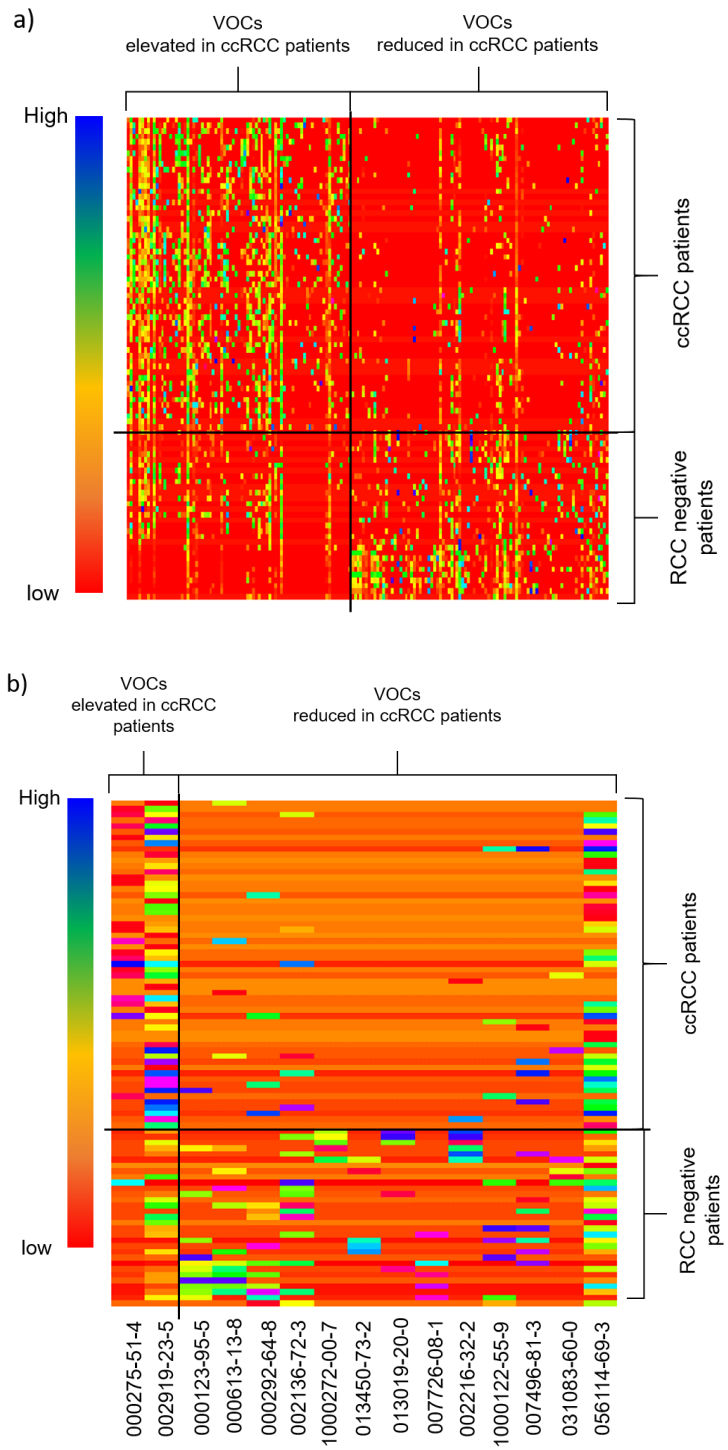


Figure 4. 1. a) The quantity heat map of significant VOCs in clear cell renal cell carcinoma (ccRCC) vs controls samples by Wilcoxon test ($p < 0.05$).; b) the quantity heat map of 15 selected VOC metabolites selected by ccRCC diagnostic model in the urines of the training cohort patients . The Chemical Abstracts Service (CAS) numbers were used to indicate each metabolite. The correlation between VOCs and patients ranges from low (red) to high (blue).

To develop the regression diagnostic model, a broader range of VOCs were selected using p=0.20. As a result, 442 potential VOCs were identified. After further selection with l_1 regularization, the final logistic model selected 15 VOCs (listed in the Table 4. 2) Among the 15 VOCs, 13 were dominated in control group and 2 dominated in cancer group as shown in Figure 4. 1b).

The performance of this ccRCC diagnostic model was validated in those 88 patients. On the basis of predicted probabilities from the final model via jackknife cross-validation, the area under the receiver operating characteristic (ROC) curve (AUC) was 0.87 with confidence interval 0.77 to 0.97 as shown in Figure 4. 2), which indicated a highly promising discrimination power between VOCs in urine of ccRCC patients and controls. The sensitivity and specificity of this ccRCC diagnostic model (normalized to Mirex) to identify patients with ccRCC was determined using ROC analysis. Compared to controls, this diagnostic model had 93% sensitivity and 77% specificity, with a cut-off point set at 0.68.

Table 4. 2. The VOCs selected by logistic regression models for clear cell renal cell carcinoma diagnosis prediction.

CAS Number[1]	Formula	Chemical Name	Dominating Group	p value[2]	Occurrence	
					Cancer(+)[3]	Control(-)[4]
000123-95-5	C22H44O2	Octadecanoic acid, butyl ester	Control	1.28E-05	1	11
000613-13-8	C14H11N	2-Anthracenamine	Control	4.54E-04	4	11
000292-64-8	C8H16	Cyclooctane	Control	1.49E-03	4	10
002136-72-3	C20H42O2	Ethanol, 2-(octadecyloxy)-	Control	1.61E-03	6	12
1000272-00-7	C9H15N3O	1,2,4-Oxadiazole, 5-methyl-3-(1-piperidylmethyl)-	Control	1.99E-03	0	5
013450-73-2	C16H17N3O	11H-Dibenzo[b,e][1,4]diazepin-11-one, 5-(3-aminopropyl)-5,10-dihydro-	Control	1.99E-03	0	5
013019-20-0	C8H16O	3-Heptanone, 2-methyl-	Control	1.99E-03	0	5
007726-08-1	C12H25NO2	Decanamide, N-(2-hydroxyethyl)-	Control	1.99E-03	0	5
002216-32-2	C9H20	Heptane, 4-ethyl-	Control	1.14E-02	2	6
1000122-55-9	C19H22N2O5	1,3-Bis-(p-carbamoylmethylphenoxy)-2-propanol	Control	3.08E-02	3	6
007496-81-3	C15H13N	Indolizine, 2-(4-methylphenyl)-	Control	4.81E-02	5	7
031083-60-0	C13H20O2	Tricyclo[4.3.1.1(3,8)]undecane-1-carboxylic acid, methyl ester	Control	8.74E-02	2	4
056114-69-3	C13H22O3Si2	Benzaldehyde, 2,5-bis[(trimethylsilyl)oxy]-	Control	1.01E-01	40	26
000275-51-4	C10H8	Azulene	Cancer	2.83E-03	18	1
002919-23-5	C4H8O	Cyclobutanol	Cancer	2.11E-02	48	23

1 Chemical Abstracts Service number

2 p value: the p value of selected compounds from Wilcoxon rank-sum Test

3 Cancer (+): clear cell renal cell carcinoma positive patient

4 Control (-): renal cancer negative patent

To evaluate the performance of the above ccRCC diagnostic model, a second cohort of patients as the testing group were recruited, and the VOCs in their urine were analyzed. Via Jackknife cross-validation, the area under the ROC curve was 0.81 with confidence interval 0.52 to 1 as shown Figure 4. 3), and had 86% sensitivity and 83% specificity (with cutoff point at 0.68). The results revealed the promising discrimination power of urinary VOCs in ccRCC diagnosis.

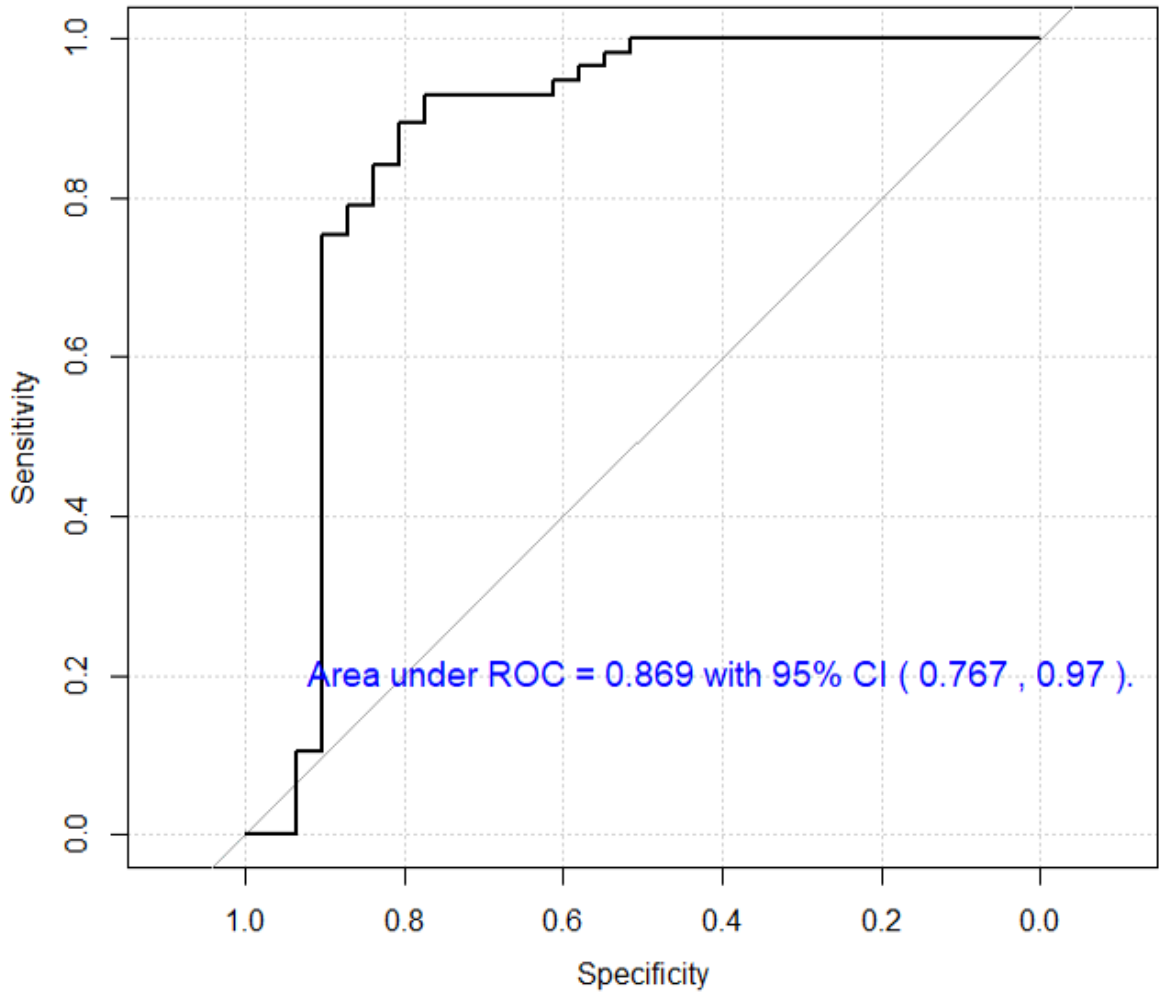


Figure 4. 2. The ROC curve for VOC ccRCC diagnosis logistic model verified in the training group with 88 patients.

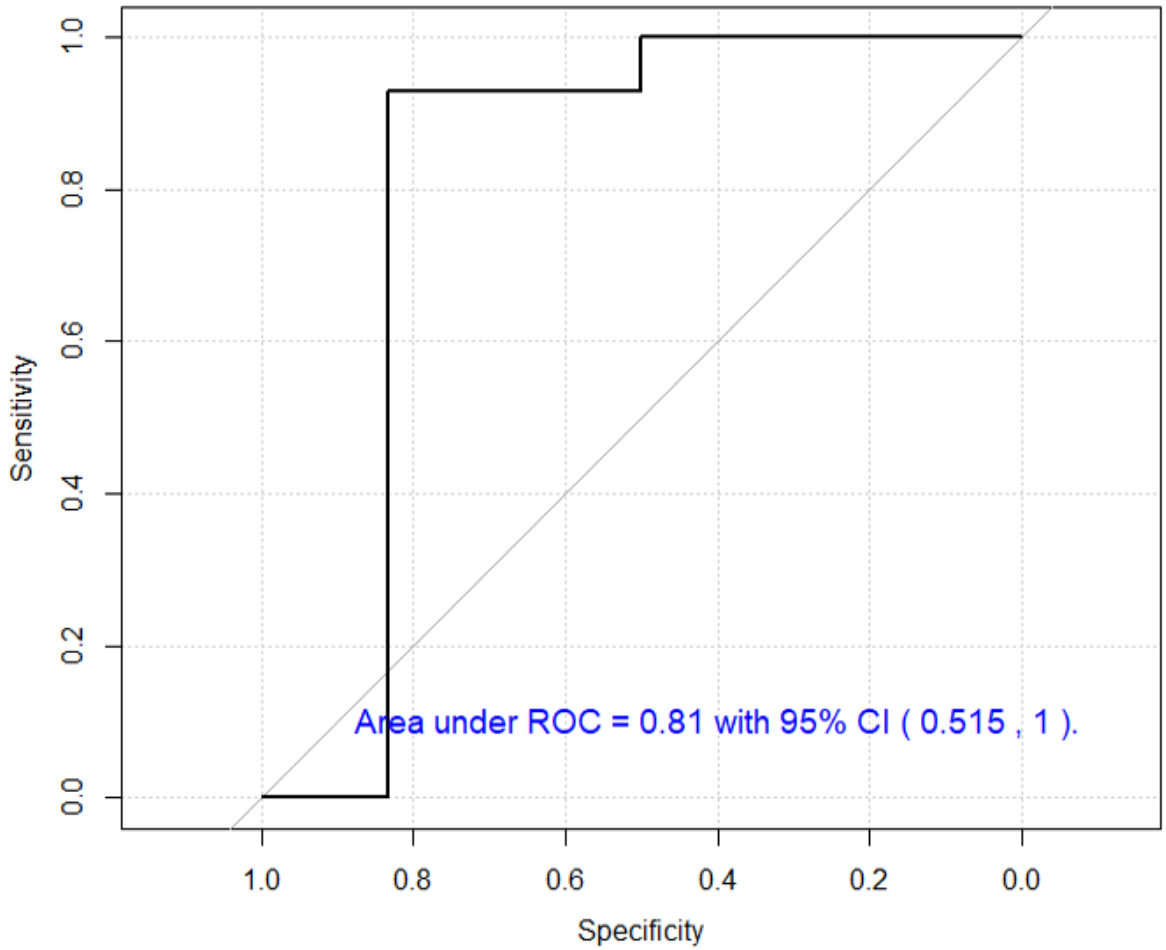


Figure 4. 3. The ROC curve for VOC ccRCC diagnosis logistic model tested in the testing group with 20 patients.

4.4 Discussion

This investigation was designed to evaluate the clinical utility of the urinary VOCs statistic model for the diagnose ccRCC. Screening through over eight thousands VOCs in urine using logistic regression, the diagnostic model including 15 VOCs was developed. The major findings in this study indicate that the VOCs based ccRCC diagnostic model had favorable sensitivity and specificity. With the area under ROC was 0.87 and 0.81, and the high sensitivity (93% and 86%) and the specificity (86% and 83%) for training data set and testing data set, respectively, the model demonstrated the clinical validity of this VOCs based ccRCC diagnostic model. To our

knowledge, it is the first model of its kind for ccRCC detection. The model offers a great potential for develop the first cost effective and non-invasive screening tool for ccRCC.

The analytical method developed in this study allowed an easy, fast and efficient analysis of VOCs without tedious sample preparation. The solventless sample preparation technique, Stir Bar Sorptive Extraction, can preserve the sample integrity and permit effective analyses for processing large sample sizes which will be an important factor for clinical translatability. Unlike gas-sensor (such as E-nose)[44], the GC/MS can provide much detailed information on the compound identity for future metabolomics and physiological studies. In this study, the VOCs detected in each sample were confirmed through the NIST library report. All samples were analyzed in a blinded and coded fashion in instrument analysis.

We also introduced an internal standard so that we were able to perform semi-quantitative analysis (i.e. relative peak area ratio to internal standard) for each VOC in the statistical examination. Mirex was selected as the internal standard because of its non-existence in human and a relative longer retention time in GC so it will not interfere with the urinary VOCs in the analysis. All the VOCs signal intensities were then normalized to that of mirex.

Unlike the determination protein biomarkers, urine aquaporin-1 (AQP1) and perilipin-2(PLIN2), of RCC reported by Jeremiah J. Morrissey et al[82], the concentration of VOCs in this study was not normalized to the concentration of urine creatinine, which is suggested to be used to minimize the impact of hydration status from patient to patient or in the same patient over time[174], because 1) there is a significant lag between the time of injury in kidney and the time that the concentration of creatinine achieve the threshold required to diagnosis kidney disease, which reveal the possible inconsistent concentration of creatinine; 2) there is no significant differences observed between creatinine normalized and non-normalized results in this protein biomarkers paper; and 3) this investigation is designed to test the performance of novel VOCs based model in ccRCC diagnosis without normalizing to any biomarker of renal disease.

Five phases of screening biomarker development were proposed Margaret Sullivan Pepe et al 2001.[174] However, our study didn't progress consecutively through those five phase. In this

investigation, the training and testing set design began with phase 2, clinical assay and validation, and phase 3, retrospective longitudinal. The phase 1, preclinical exploratory, are being investigated along with the phase 2 and phase 3 in this study. Furthermore, even though the two urinary exosomal proteins, AQP-1 and PLIN2, have shown promise as the biomarkers in RCC diagnosis, they can be found in clear cell and papillary RCC but not in the chromophobe subtype of RCC.[82] Our VOCs based screening model, however, has great potential to be developed as a more universal screening tool of almost all types of RCC because of the metabolic distinction shown between cancer patients and controls. Unlike the ELISA detection methods of AQP-1 and PLIN2, the VOCs based diagnostic model is a high throughput and fast screening method in clinical applications enabled by high performance SBSE, GC/MS and statistical analysis.

The potential of urinary VOCs used in RCC diagnose has been highlighted in previous studies.[22,80,81] Most of the studies focused on the search of specific VOCs in RCC patients without further validation.[22,80] In the study reported by Marica Monteiro in 2017[81], the selected VOCs was validated in different patients group for the existence, but the performance (discrimination power) of VOCs in differentiating RCC patients and controls was not determined. Our study, however, have shown the prediction power of the VOCs based diagnostic model which was determined through the statistical analysis with AUC, sensitivity, specificity, and accuracy calculation. Another major difference is the introduction of internal standard, Mirex, which enables the relative more accurate quantity determination of VOCs in urine.

We further studied the VOCs selected for the ccRCC screening model. When examining their chemical structures, Octadecanoic acid-butyl ester, cyclooctane, 2-(octadecyloxy)-ethanol, 2-methyl-3-heptanone, cyclobutanol, 4-ethyl-heptane, and N-(2-hydroxyethyl)-decanamide may be involved in peroxidation of fatty acids because of the long carbon chain and/or carbanyl group. Octadecanoic acid-butyl ester and 2-(octadecyloxy)-ethanol could be the peroxidation products of oleic acid ($C_{18}H_{34}O_2$), which is one of the main (poly) unsaturated fatty acids (P)UFA *in vivo* and has potent anti-inflammatory properties.[175,176] In the structure of N-(2-hydroxyethyl)-decanamide, a compound in control group, the moiety of ethanolamine can be formed through the

decarboxylation of serine or the reduction of glycine.[177] Ethanolamine, serine, glycine and sarcosine are closely interrelated in several one-carbon metabolic cycles. One-carbon metabolism integrates cellular nutrient status by cycling carbon units from amino acid inputs to generate diverse outputs, including redox maintenance and cellular biosynthesis [99]. Genetic and functional evidence suggests that hyper activation of those pathways is a driver of oncogenesis and establishes a link to cellular epigenetic status [99]. Additionally, the plasma membrane of normal cells is characterized by an asymmetric distribution of various phospholipids over two membrane leaflet. Phosphatidylethanolamine (PE) resides in the inner leaflet facing the cytosol. The disrupted membrane asymmetry of cancer cell exposes PE to extracellular space [124], and the increasing need of PE in cancer cell may correlate the excessive consumption of ethanolamine (which was found to be mostly appear in control group in our study) and enhanced lipogenesis.

This study had cover three of the five phases of screening biomarker development proposed by Margaret Sullivan Pepe et al 2001.[174] In this study, the training and testing set design began with phase 2, clinical assay and validation, and phase 3, retrospective longitudinal. The phase 1, preclinical exploratory, are being investigated along with the phase 2 and phase 3 in this study. One of the aims in this investigation was to test the value of VOCs based ccRCC diagnostic model to screen and identify ccRCC. Even with a considerably small sample size, the VOCs based model showed very promising discriminating power between ccRCC and control patients. A future study could include larger cohort of patients and expand to other types of RCC.

4.5 Summary

This investigation was designed to evaluate the clinical utility of the urinary VOCs statistic model for the diagnose ccRCC. The findings indicated that the VOCs based ccRCC diagnostic model had favorable sensitivity and specificity. With the area under ROC was 0.87 and 0.81, and the high sensitivity (93% and 86%) and the specificity (86% and 83%) for training data set and

testing data set, respectively, This investigation supports the ability of urinary VOCs based diagnostic model to early and non-invasive screening ccRCC patients.

Overall, it validates the clinical utility of urinary VOCs based diagnostic model as the biomarker for ccRCC. In addition, the validation of those selected VOCs in a larger cohort would confirm the statistical significance in ccRCC screening and even in ccRCC staging diagnose.

Chapter 5: Cross Examination of Volatile Organic Compounds (VOCs) based models in prostate cancer and clear cell renal carcinoma diagnosis

- Two urinary VOCs based models were independently developed previously for prostate cancer (PCa) and clear cell renal cell carcinoma (ccRCC). This chapter was to cross examine the VOC-PCa model on ccRCC and the ccRCC VOC model on PCa detection, to study the model cancer specificity for the two promising cancer diagnostic models.
- The area under received operation curves from the validations verifies the cancer specific discrimination power of the VOCs models for PCa and ccRCC screening.

Abstract:

Background: In our previous studies, the VOCs based diagnosis models were respectively developed and validated in different cohort of patients to differentiate PCa and ccRCC patients. To test the performance of the models with increased complexity, we reevaluated the performance of those two VOCs models.

Methods: Two previously developed urinary VOCs based models were included in this study: PCa model (11 VOCs included) and ccRCC model (15 VOCs included). Four patient groups involved in this study were 55 PCa positive patients, 53 PCa negative patients, 55 ccRCC positive patients and 31 RCC negative patients as a larger cohort. The performance of PCa model to differentiate PCa positive patients from others and ccRCC model to differentiate ccRCC positive patients from others were evaluated via the Receiver Operating Characteristic (ROC) curve.

Results: The discrimination power of PCa model resulted in an area under ROC curve (AUC) of 0.834 with confidence interval of 0.779 to 0.889. Then, after testing the ccRCC model, the AUC was 0.779 with confidence interval 0.707 to 0.851.

Conclusion: The study verified the discrimination power of the PCa model and ccRCC model with increased population complexity.

Clinical Practice Points

- The internal validation results of two urinary VOCs based models, PCa model and ccRCC model, were cross examined on patients who have different types of cancer. The patient complexities in clinical practice and the specificity of those cancer diagnostic models should be taken into consideration.
- We cross examined the PCa model and ccRCC in a larger sample pool with increased patient complexity. The results from the logistic regression showed good outcomes (i.e. high AUC), which verified the reliability of VOCs based cancer diagnostic models for prostate cancer and clear cell renal carcinoma diagnosis.
- As a non-invasive and sensitive tool, the VOC model is ideally suited for early diagnosis of PCa and ccRCC. Future validation of our models will enable clinically relevant and cost-effective, point-of-care PCa, ccRCC and even other urinary tract cancers screening and prognosis with the potential to save health care costs.

5.1 Introduction

In our previous studies, the VOCs based diagnostic models were developed and validated in different cohorts of patients to differentiate prostate cancer (PCa), and clear cell renal cell carcinoma (ccRCC) patients from controls patients. The increasing availability and use of predictive models highlights the need for careful assessment of the validity of these models.[178] Meanwhile, the challenges of validating a biomarker-based predictive model are considerable, and the validation process should be planned as early as possible in the design phase of risk prediction model development.

Validation requires a comparison of model prediction with a "gold standard", i.e. the true outcome.[178] The "gold standard" must be ascertainable or observable in the future and might be able to indicate whether the patient has cancer or will they eventually have a recurred cancer confirmed by a procedure (such as biopsy) or long term follow-up. For example, the true outcome

in the Prostate Cancer Prevention Trial (PCPT) is prostate cancer status obtained via biopsy.[160] The input data, prediction, and all the other available truth for the patients are all required in the validation. Furthermore, the larger validation data sets are preferable and crucial in the evaluation of a risk model. In general, there are two forms of validation, internal validation and external validation.[178] Internal validation involves training-testing splits of the available data and/or cross validation, as designed in our previous studies (chapter 2 and 4).[38] It can provide the valid assessment of model performance in disease risk prediction. External validation are conducted through a different data set collected by different investigators from different institutions. Validation on heterogeneous external data sets allow for the evaluation of predictive model on populations other than the one on which the model was developed.[178] For example, different geographic locations or different clinical practices would contribute to the difference of the populations.

The internal validation results from our research have revealed the great potential of those VOCs based models to facilitate informed decision making in clinical practice. However, the population complexities in clinical practice and the specificity of those cancer diagnostic models should be taken into consideration. The VOCs based diagnostic models were developed based on the patients who were confirmed with particular cancer (either prostate cancer or renal cancer) by clinical procedures, and those models are expected to be used in early screening and diagnosis of PCa and ccRCC. The ability of those VOCs based models to differentiate PCa or ccRCC patients from other urinary tract cancers is unknown. Therefore, the objective of this investigation was to conduct a "pre-external validation" and cross examine the discrimination performance of the PCa and ccRCC model among the patient groups who either have different type of cancer or other medical conditions not related to the disease of the model. In this study, we hypothesized that the VOCs found in either PCa or ccRCC are cancer specific. Therefore the cancer specific VOC models would be able to differentiate PCa or ccRCC cancer patients from those who do not have these specific cancer regardless of the possible existence of other cancer.

5.2 Experimental Section

5.2.1 Study Design

This study was aimed to conduct the "pre-external validation" and reevaluate the discrimination performance of the above VOCs based PCa model and ccRCC model for the cancer type specificity and larger sample validation. PCa model was tested between PCa patients and those who do not have PCa, e.g. ccRCC patients plus all the controls; and the ccRCC model was tested between ccRCC and those who do not have ccRCC, e.g. PCa patients plus all the controls (Figure 5.1).

Patient recruitment and approval procedures followed the same procedures as described in chapter 2 and 4. Briefly, approval was obtained from the Internal Review Board, and written informed consent was obtained from all patients. Both PCa Patients presented to Geisinger Medical Center urologic clinic for evaluation of elevated PSA >2.5 ng/ml or abnormal digital rectal exam and their cancer status was confirmed by prostate biopsy. As for ccRCC groups, cancer status were confirmed by CT imaging procedure. All patients provided urine specimens for dipstick urinalysis and the remaining urine samples (50 ml) were collected and stored at -80^{o}C prior to VOC analysis. In this study, 55 diagnosed with PCa, while 53 were negative for PCa; 57 patients with ccRCC and 31 patients RCC negative based the imaged renal mass. The demographic information of all those patients are shown in Table 5. 1.

Table 5. 1 Demographic information of prostate cancer, clear cell renal carcinoma, and control patients in the reevaluation of VOCs based cancer diagnostic models. Data are presented as median (interquartile range) for continuous variables and n (%) for categorical variables.

	Prostate cancer patients		
	Cancer patients	Controls	p value [b]
N	55	53	
PSA [a] (ng/mL)	5.29 (0.08-1987)	2.6 (0.1-18.2)	0.28
Gleason score		N/A	
6	20 (36%)		
7	23 (42%)		
8	6 (11%)		
9	6 (11%)		
	Renal cancer patients		
	Cancer patients	Controls	
N	57	31	
Tumor grade [c]		N/A	
1	3 (5%)		
2	24 (42%)		
3	18 (32%)		
4	12 (21%)		

a. PSA: prostate specific antigen
b. The p value from the t-test of the PSA numbers between prostate cancer and control groups
c. Pathology/Biopsy Confirmed Dx (Tumor grade

The validation procedures in this study will be divided into two parts shown in as shown in Figure 5. 1 1) PCa diagnostic model would be validated in the PCa (+) and all other three groups (i.e. ccRCC patients, ccRCC (-) and PCa (-); 2) ccRCC diagnostic model was validated in the ccRCC and all other ccRCC negative group (i.e. ccRCC (-), PCa (+) and PCa (-).

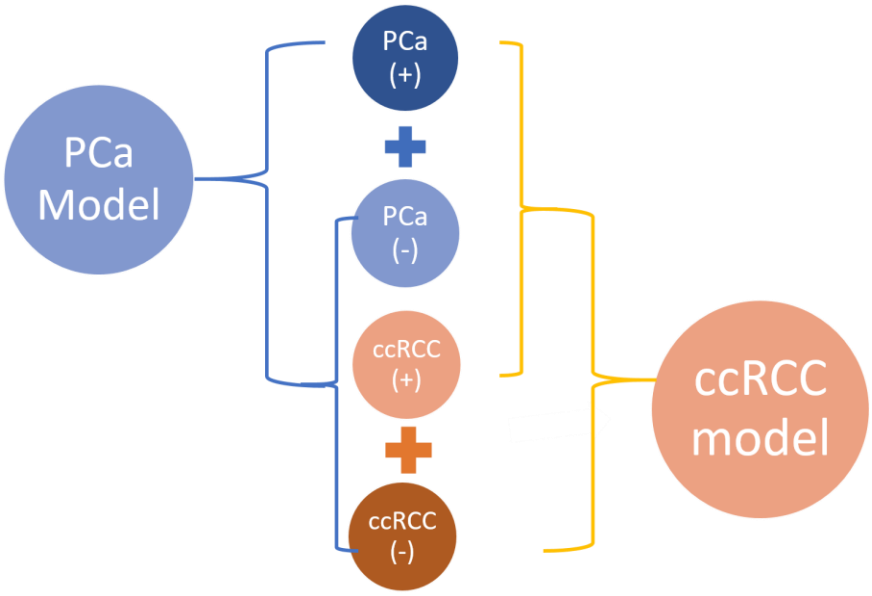


Figure 5. 1 The study design of reevaluation of 11 VOCs based PCa diagnostic model and 15 VOCs based ccRCC diagnostic model

5.2.2 Data processing and statistical analysis

The cross examination of PCa and ccRCC logistic models were evaluated via the Receiver Operating Characteristic (ROC) curve and its associated performance was measured on the basis of its jackknife prediction (Kleinbaum and Klein, 2010 [138]). All the analyses are performed using the open-source statistical computing package R[141]. The details were previously published in by Gao et al.[38]

5.3 Results

There were 11 VOCs included in the PCa logistic model while the ccRCC logistic model selected 15 VOCs. As shown in .

Table 5. 2 there is no duplicated VOCs between those two groups: these 11 VOCs for PCa diagnosis and 15 VOCs for ccRCC diagnosis.

Table 5. 2 11 VOCs selected in prostate cancer diagnostic models and 15 VOCs selected in clear cell renal carcinoma diagnostic models.

11 VOCs PCa model		15 VOCs ccRCC model	
CAS Number	**Chemical name**	**CAS Number**	**Chemical name**
000472-41-3	4-(3,4-dihydro-2,2,4-trimethyl-2H-1-benzopyran-4-yl)-phenol	000123-95-5	Octadecanoic acid, butyl ester
000050-28-2	Estradiol	000613-13-8	2-Anthracenamine
129086-73-3	Ethyl à-hydroxymyristate trisiloxane	000292-64-8	Cyclooctane
1000215-25-2	1-(2,4-Dimethylphenyl)-3-(tetrahydrofuryl-2)propane	002136-72-3	Ethanol, 2-(octadecyloxy)-
1000126-50-5	2-amino-Imidazole-5-carboxylic acid	1000272-00-7	1,2,4-Oxadiazole, 5-methyl-3-(1-piperidylmethyl)-
000995-83-5	1,1,3,3,5,5,7,7,9,9-decamethyl-pentasiloxane	013450-73-2	11H-Dibenzo[b,e][1,4]diazepin-11-one, 5-(3-aminopropyl)-5,10-dihydro-
003555-47-3	1,1,1,5,5,5-hexamethyl-3,3-bis[(trimethylsilyl)oxy]-Trisiloxane	013019-20-0	3-Heptanone, 2-methyl-
020548-62-3	Phthalic acid, bis(7-methyloctyl) ester	007726-08-1	Decanamide, N-(2-hydroxyethyl)-
024535-53-3	4-Nitro-4'-chlorodiphenylsulfoxide	002216-32-2	Heptane, 4-ethyl-
075132-80-8	1-Propylpentachlorotriphosphazene	1000122-55-9	1,3-Bis-(p-carbamoylmethylphenoxy)-2-propanol
101100-38-3	2,6-di-t-butyl-4-hydroxymethylene-2,3,5,6-detetrahydrocyclohexanone	007496-81-3	Indolizine, 2-(4-methylphenyl)-
		031083-60-0	Tricyclo[4.3.1.1(3,8)]undecane-1-carboxylic acid, methyl ester
		056114-69-3	Benzaldehyde, 2,5-bis[(trimethylsilyl)oxy]-
		000275-51-4	Azulene
		002919-23-5	Cyclobutanol

The PCa model was evaluated to differentiate PCa patients from ccRCC patients and all controls (i.e. PCa (-) and ccRCC (-). Via Jackknife cross-validation, the area under the ROC curve is 0.834 with confidence interval 0.779 to 0.889 as shown in Figure 5. 2, which verified the promising discrimination power of urinary VOCs in PCa diagnosis and revealed the prostate cancer specificity of this 11 VOCs model.

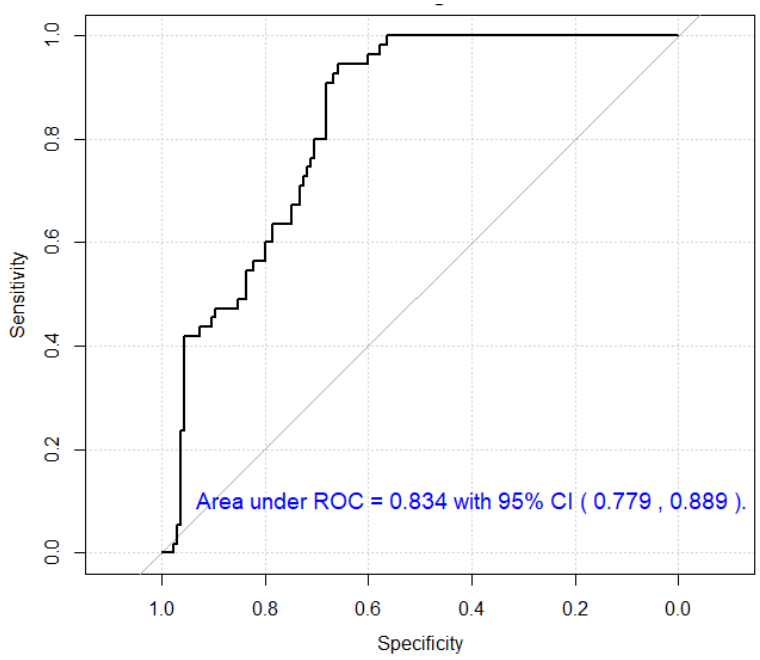


Figure 5. 2 The ROC curve for VOC PCa diagnosis logistic model tested in 196 patients.

Then, ccRCC model was validated in discrimination ccRCC from PCa patients and all controls (i.e. PCa (-) and ccRCC (-). Through the same cross-validation method, the AUC is 0.779 with confidence interval 0.707 to 0.851 as shown in Figure 5. 3, which proved the potential of urinary VOCs in ccRCC diagnosis.

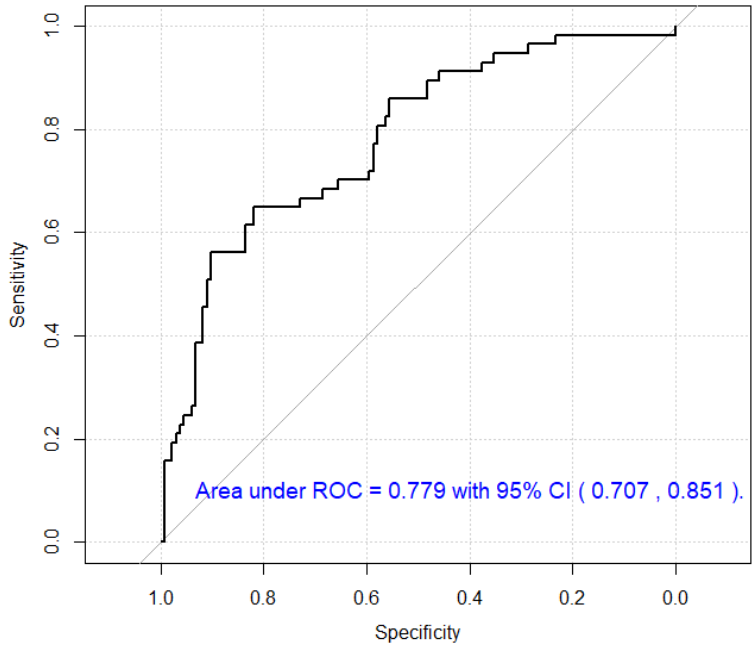


Figure 5. 3 The ROC curve for VOC ccRCC diagnosis logistic model tested in 196 patients.

5.4 Discussion

The integrity and reliability of disease prediction models could be secured through careful assessment of the validity of these models.[178] There are two reasons to involve the careful validation for the model developed from disease biomarkers, 1) overfitting issues with complex models involving a large number of biomarkers, and 2) inter laboratory variation in assays used to measure biomarkers.[178] Those are also the reasons that we involved training-testing validation method to evaluate the VOCs models for PCa diagnosis[38] and ccRCC diagnosis and conducted the cross examination of those two models with complexed patient groups. In this study, we cross examined the performance of the PCa model predication and ccRCC prediction model with different sample pool which presented a more complex medical scenarios. In the cross evaluation of the PCa model, PCa negative group also contained patients who either have ccRCC or presented to the facility with other kidney problem. Likewise, the ccRCC was evaluated against patients who appeared to have prostate problems. The AUC 0.834 and 0.779 for validation sample pool

with increased complexity has verified the potential of the VOCs model in PCa diagnosis and the VOCs model in ccRCC diagnosis again. The results revealed the promising discrimination power of those two VOCs model to differentiate the targeted cancer from a complex groups, which provided a promising outlook of external validation of those VOCs based models in the near future.

A guidance of five phases in biomarker development for early detection of cancers were proposed by Pepe et al[179]. As the investigation of VOCs based cancer prediction models in our studies were conducted through mass spectrometry detection and innovative statistical machine-learning techniques, the development of VOCs models for PCa and ccRCC prediction could have easily passed phase 1 (identify promising directions) and showed favorable results in phase 2 (clinical assay detects established disease). The further phase 3 (retrospective longitudinal repository studies), phase 4 (prospective screening studies) and phase 5 (cancer control studies) would be further pursued in extended internal validation and external validation of our studies.

5.5 Summary

This investigation supports the ability of urinary VOCs based diagnostic model to early and non-invasive screening PCa and ccRCC patients. It further validates the clinical utility of urinary VOCs based diagnostic model as the biomarker for PCa and ccRCC. In addition, the external validation of those selected VOCs of PCa prediction model and ccRCC model in a larger cohort would confirm the statistical significance in PCa and ccRCC screening.

Chapter 6: Conclusions and Outlooks

- This chapter states the concluding remarks and future directions of the research work in this dissertation.

6.1 Concluding remarks

As the metabolites produced by human body, urinary volatile organic compounds (VOCs) of cancer patients and controls were investigated to develop VOCs based models for prostate cancer (PCa) and clear cell renal carcinoma (ccRCC) diagnosis. Collaborating with clinical researchers and experts in statistics, we utilized advanced analytical chemistry to develop innovative machine-learning VOCs based model for cancer early detection. The ease of sample preparation (i.e. Stir Bar Sorptive Extraction) and sensitive analytical instrumentation (Gas Chromatography-Mass Spectrometry) allowed high throughput of VOC detection and that enabled the research to unbiasedly search for significant VOCs for cancer early detection. Two urinary VOCs based diagnostic models were developed: PCa model and ccRCC model. The development, validation and even cross examination of these models verified the great discrimination power of urinary VOCs for differentiating cancer patients from controls, and high/intermediate risk patients from low risk patients. The performance of these models were summarized in Table 6. 1. The PCa model was based on 11 significant VOCs; while the ccRCC diagnosis model involved 15 VOCs In addition, we were able to show that VOCs based model could be used for PCa risk assessment.

Table 6. 1 The summary of VOCs based cancer diagnosis models for prostate cancer and renal cancer

VOCs Disease Models	PCa Model	PCa Risk Assessment	ccRCC Model		
# of Dominated VOCs ($p < 0.05$)				**# of VOCs overlapped between PCa and PCa risk assessment**	**# of VOCs overlapped between PCa and ccRCC**
Elevated in Cancer (or in High/Intermediate risk)	254	23	79	5	10
Reduced in Cancer (or in High/Intermediate risk)	282	44	91	1	7
VOCs selected in final model					
# of VOCs	11	11	15	0	0
AUC of training	0.92	0.86	0.87		
AUC of testing	0.86		0.81		
Sensitivity	96%	85%	93%		
Specificity	80%	80%	77%		
cross examination AUC in complex sample pool	0.83		0.78		

The urinary VOC based metabolites approach could likely be adapted into a clinically viable, highly sensitive, cost-effective portable diagnostic assay for PCa, ccRCC and even other cancers. Nonetheless, the validity and effectiveness of those selected VOCs in PCa and ccRCC diagnosis and PCa risk assessment need to be further confirmed in a larger cohort study.

6.2 Future directions

6.2.1 Larger cohort validation

The validity and effectiveness of those selected VOCs in PCa and ccRCC diagnosis and PCa risk assessment need to be further confirmed in a larger cohort study. At the same time, there are some further aims that should be included in the larger cohort evaluation of those VOCs based cancer diagnosis models. Those include:

1) to build up risk score system for prostate cancer and renal cancer prediction;
2) to integrate urine sample screening system including VOCs profile detection, statistical risk prediction and report production;
3) to estimate the reductions in cancer mortality and over-diagnosis of prostate cancer and renal cancer which could be afforded by the VOCs based diagnosis models; and
4) to study factors (such as age, ethnicity, dilution factor, etc.) that could affect the prediction ability of the models in cancer detection.

As VOCs reflect the physiological and metabolic status of the individual, they could be further investigated in basic prostate cancer metabolomics research. These VOCs, that are only produced in cancer cells or produced with significantly higher or lower levels than normal cells, may therefore serve as biomarkers for the assessment, detection, monitoring, and even treatment of disease. Though the study of the pathways affecting VOCs production is not within the scope of this project, it is our goal to confirm those VOCs that are characteristic of cancers so that the finding can be used in exploring pathway studies.

6.2.2 Model validation with samples with/without treatment

The comparison about the behavior of VOCs before and after different treatments of prostate cancer and renal caner is still unknown. And it would be the one of future direction of VOCs research as cancer biomarker. In fact, the progression of cancers is different from each other and the urinary VOCs profiles could play different roles in different types and stage of cancers. Prostate cancer is a heterogeneous disease ranging from indolent to life threatening stages. The monitoring of urinary VOCs profile with patients with localized tumor would be a reliable, fast, low-cost and patients-friendly alternative other than invasive PSA serum test and DRE. For renal cancer, the same type monitoring of urine VOCs profile could help those patients with kidney disease, who may have higher incidence of renal cancer. Then, the track of urinary VOCs profile after surgery and/or along the treatment, like the chemotherapy, would be helpful in early detection of cancer recurrence and even the second cancer.

In the United States, prostate cancer (PCa) is the most common cancer and the third leading cause of death in men. Routine PSA testing was discouraged by the United Service Preventive Services Task Force in 2008 due to conflicting data questioning its value in reducing PCa-specific mortality. Kidney cancer (KCa) accounts for more than 2% of cancer incidence and mortality in the United States while there is no early detection method available. Our goal is to develop a cost effective, accurate, and non-invasive cancer screening methods, and provide doctors alternatives for more accurate diagnosis, and reduce cost in the healthcare system from ambiguous and unnecessary recommendations for biopsy.

References

1 Gallagher, M. *et al.* Analyses of volatile organic compounds from human skin. *British Journal of Dermatology* **159**, 780-791 (2008).

2 Ashley, D. L. *et al.* Determining volatile organic compounds in human blood from a large sample population by using purge and trap gas chromatography/mass spectrometry. *Analytical chemistry* **64**, 1021-1029 (1992).

3 Phillips, M. *et al.* Variation in volatile organic compounds in the breath of normal humans. *Journal of Chromatography B: Biomedical Sciences and Applications* **729**, 75-88 (1999).

4 de Lacy Costello, B., Ratcliffe, N. M. & Smith, D. *Volatile organic compounds (VOCs) found in urine and stool*. (Elsevier, Amsterdam, The Netherlands, 2013).

5 Amann, A. & Smith, D. *Volatile biomarkers: non-invasive diagnosis in physiology and medicine*. (Newnes, 2013).

6 Shirasu, M. & Touhara, K. The scent of disease: volatile organic compounds of the human body related to disease and disorder. *The Journal of Biochemistry* **150**, 257-266 (2011).

7 Balseiro, S. & Correia, H. Is olfactory detection of human cancer by dogs based on major histocompatibility complex-dependent odour components?–A possible cure and a precocious diagnosis of cancer. *Medical hypotheses* **66**, 270-272 (2006).

8 Cornu, J.-N., Cancel-Tassin, G., Ondet, V., Girardet, C. & Cussenot, O. Olfactory detection of prostate cancer by dogs sniffing urine: a step forward in early diagnosis. *European urology* **59**, 197-201 (2011).

9 Boedeker, E., Friedel, G. & Walles, T. Sniffer dogs as part of a bimodal bionic research approach to develop a lung cancer screening. *Interactive cardiovascular and thoracic surgery* **14**, 511-515 (2012).

10 Willis, C. M., Britton, L. E., Harris, R., Wallace, J. & Guest, C. M. Volatile organic compounds as biomarkers of bladder cancer: Sensitivity and specificity using trained sniffer dogs. *Cancer Biomarkers* **8**, 145-153 (2011).

11 Horvath, G., Andersson, H. & Paulsson, G. Characteristic odour in the blood reveals ovarian carcinoma. *BMC cancer* **10**, 643 (2010).

12 Wang, C. *et al.* Blood volatile compounds as biomarkers for colorectal cancer. *Cancer biology & therapy* **15**, 200-206 (2014).

13 Deng, C., Zhang, X. & Li, N. Investigation of volatile biomarkers in lung cancer blood using solid-phase microextraction and capillary gas chromatography–mass spectrometry. *Journal of Chromatography B* **808**, 269-277 (2004).

14 Goldberg, E., Blendis, L. & Sandler, S. A gas chromatographic—mass spectrometric study of profiles of volatile metabolites in hepatic encephalopathy. *Journal of Chromatography B: Biomedical Sciences and Applications* **226**, 291-299 (1981).

15 Manolis, A. The diagnostic potential of breath analysis. *Clinical chemistry* **29**, 5-15 (1983).

16 Phillips, M. *et al.* Variation in volatile organic compounds in the breath of normal humans. *Journal of Chromatography B: Biomedical Sciences and Applications* **729**, 75-88 (1999).

17 Phillips, M. *et al.* Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *The Lancet* **353**, 1930-1933 (1999).

18 Peng, G. *et al.* Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nature nanotechnology* **4**, 669 (2009).

19 Banday, K. M. *et al.* Use of urine volatile organic compounds to discriminate tuberculosis patients from healthy subjects. *Analytical chemistry* **83**, 5526-5534 (2011).

20 Sethi, S., Nanda, R. & Chakraborty, T. Clinical application of volatile organic compound analysis for detecting infectious diseases. *Clinical microbiology reviews* **26**, 462-475 (2013).
21 Khalid, T. *et al.* Urinary volatile organic compounds for the detection of prostate cancer. *PloS one* **10**, e0143283 (2015).
22 Monteiro, M. *et al.* Analysis of volatile human urinary metabolome by solid-phase microextraction in combination with gas chromatography–mass spectrometry for biomarker discovery: application in a pilot study to discriminate patients with renal cell carcinoma. *European Journal of Cancer* **50**, 1993-2002 (2014).
23 Weber, C. M. *et al.* Evaluation of a gas sensor array and pattern recognition for the identification of bladder cancer from urine headspace. *Analyst* **136**, 359-364 (2011).
24 Calderón-Santiago, M. *et al.* Human sweat metabolomics for lung cancer screening. *Analytical and bioanalytical chemistry* **407**, 5381-5392 (2015).
25 Choi, M.-J. & Oh, C.-H. 2nd dimensional GC-MS analysis of sweat volatile organic compounds prepared by solid phase micro-extraction. *Technology and Health Care* **22**, 481-488 (2014).
26 Garner, C. E., Smith, S., Bardhan, P., Ratcliffe, N. M. & Probert, C. A pilot study of faecal volatile organic compounds in faeces from cholera patients in Bangladesh to determine their utility in disease diagnosis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**, 1171-1173 (2009).
27 Garner, C. E. *et al.* Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *The FASEB Journal* **21**, 1675-1688 (2007).
28 Batty, C. A., Cauchi, M., Lourenco, C., Hunter, J. O. & Turner, C. Use of the analysis of the volatile faecal metabolome in screening for colorectal cancer. *PLoS One* **10**, e0130301 (2015).
29 Schmidt, K. & Podmore, I. Current challenges in volatile organic compounds analysis as potential biomarkers of cancer. *Journal of biomarkers* **2015** (2015).
30 Poli, D. *et al.* Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatisation SPME–GC/MS. *Journal of Chromatography B* **878**, 2643-2651 (2010).
31 Arthur, C. L. & Pawliszyn, J. Solid phase microextraction with thermal desorption using fused silica optical fibers. *Analytical chemistry* **62**, 2145-2148 (1990).
32 Zhang, Z. & Pawliszyn, J. Headspace solid-phase microextraction. *Analytical chemistry* **65**, 1843-1852 (1993).
33 Pawliszyn, J. *Solid phase microextraction: theory and practice*. (John Wiley & Sons, 1997).
34 Pawliszyn, J. *Handbook of solid phase microextraction*. (Elsevier, 2011).
35 Baltussen, E., Sandra, P., David, F. & Cramers, C. Stir bar sorptive extraction (SBSE), a novel extraction technique for aqueous samples: theory and principles. *Journal of Microcolumn Separations* **11**, 737-747 (1999).
36 Melo, L., Nogueira, A., Lancas, F. & Queiroz, M. Polydimethylsiloxane/polypyrrole stir bar sorptive extraction and liquid chromatography (SBSE/LC-UV) analysis of antidepressants in plasma samples. *Analytica chimica acta* **633**, 57-64 (2009).
37 Soini, H., Bruce, K., Wiesler, D. & Novotny, M. Quantification of Volatiles in Mammalian Urine by Stir Bar Sorptive Extraction (SBSE) Techniques and Gas Chromatography.

38 Gao, Q. *et al.* Application of Urinary Volatile Organic Compounds (VOCs) for the diagnosis of Prostate Cancer. *Clinical Genitourinary Cancer* (2019).
39 Tienpont, B., David, F., Bicchi, C. & Sandra, P. High capacity headspace sorptive extraction. *Journal of Microcolumn Separations* **12**, 577-584 (2000).
40 Ochiai, N. *et al.* Determination of trace amounts of off-flavor compounds in drinking water by stir bar sorptive extraction and thermal desorption GC-MS. *Analyst* **126**, 1652-1657 (2001).
41 Phillips, M. Method for the collection and assay of volatile organic compounds in breath. *Analytical biochemistry* **247**, 272-278 (1997).
42 Li, N. *et al.* Gas chromatography–mass spectrometric analysis of hexanal and heptanal in human blood by headspace single-drop microextraction with droplet derivatization. *Analytical biochemistry* **342**, 318-326 (2005).
43 Peng, G. *et al.* Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nature nanotechnology* **4**, 669-673 (2009).
44 Nakhleh, M. K. *et al.* Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules. *ACS nano* (2016).
45 Filipiak, W. *et al.* TD-GC-MS analysis of volatile metabolites of human lung cancer and normal cells in vitro. *Cancer Epidemiology and Prevention Biomarkers* **19**, 182-195 (2010).
46 Fuchs, P., Loeseken, C., Schubert, J. K. & Miekisch, W. Breath gas aldehydes as biomarkers of lung cancer. *International Journal of Cancer* **126**, 2663-2670 (2010).
47 Ligor, T. *et al.* The analysis of healthy volunteers' exhaled breath by the use of solid-phase microextraction and GC-MS. *Journal of breath research* **2**, 046006 (2008).
48 Wehinger, A. *et al.* Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas. *International Journal of Mass Spectrometry* **265**, 49-59 (2007).
49 Smith, D. & Španěl, P. Selected ion flow tube mass spectrometry (SIFT‐MS) for on‐line trace gas analysis. *Mass spectrometry reviews* **24**, 661-700 (2005).
50 Westhoff, M. *et al.* Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of patients with lung cancer: results of a pilot study. *Thorax* **64**, 744-748 (2009).
51 Di Natale, C. *et al.* Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. *Biosensors and Bioelectronics* **18**, 1209-1218 (2003).
52 Hagleitner, C. *et al.* Smart single-chip gas sensor microsystem. *Nature* **414**, 293 (2001).
53 Di Francesco, F., Fuoco, R., Trivella, M. G. & Ceccarini, A. Breath analysis: trends in techniques and clinical applications. *Microchemical journal* **79**, 405-410 (2005).
54 Oh, E. H., Song, H. S. & Park, T. H. Recent advances in electronic and bioelectronic noses and their biomedical applications. *Enzyme and microbial technology* **48**, 427-437 (2011).
55 Biasioli, F., Yeretzian, C., Märk, T. D., Dewulf, J. & Van Langenhove, H. Direct-injection mass spectrometry adds the time dimension to (B) VOC analysis. *TrAC Trends in Analytical Chemistry* **30**, 1003-1017 (2011).
56 Röck, F., Barsan, N. & Weimar, U. Electronic nose: current status and future trends. *Chemical reviews* **108**, 705-725 (2008).
57 World Health Organization. CANCER CONTROL: A GLOBAL SNAPSHOT IN 2015. World Health Organization; 2016.

58 Wood, S. L., Knowles, M. A., Thompson, D., Selby, P. J. & Banks, R. E. Proteomic studies of urinary biomarkers for prostate, bladder and kidney cancers. *Nature Reviews Urology* **10**, 206 (2013).
59 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA: a cancer journal for clinicians* **68**, 7-30 (2018).
60 Richter, J. *et al.* High-throughput tissue microarray analysis of cyclin E gene amplification and overexpression in urinary bladder cancer. *The American journal of pathology* **157**, 787-794 (2000).
61 Rogers, M. A. *et al.* Proteomic profiling of urinary proteins in renal cancer by surface enhanced laser desorption ionization and neural-network analysis: identification of key issues affecting potential clinical utility. *Cancer research* **63**, 6971-6983 (2003).
62 Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910 (2009).
63 American Cancer Society. Tests for Prostate Cancer. Access at www.cancer.org/cancer/prostate-cancer/detection-diagnosis-staging/how-diagnosed.html#references on Mar 11, 2016.
64 Moyer, V. A. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. *Annals of internal medicine* **157**, 120-134 (2012).
65 Schröder, F. H. *et al.* Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine* **360**, 1320-1328 (2009).
66 Parekh, D. J. *et al.* A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. *European urology* **68**, 464-470 (2015).
67 Leyten, G. H. *et al.* Prospective multicentre evaluation of PCA3 and TMPRSS2-ERG gene fusions as diagnostic and prognostic urinary biomarkers for prostate cancer. *European urology* **65**, 534-542 (2014).
68 Chun, F. K. *et al.* Prostate cancer gene 3 (PCA3): development and internal validation of a novel biopsy nomogram. *European urology* **56**, 659-668 (2009).
69 Loeb, S. & Catalona, W. J. The Prostate Health Index: a new test for the detection of prostate cancer. *Therapeutic advances in urology* **6**, 74-77 (2014).
70 Wojno, K. J. *et al.* Reduced rate of repeated prostate biopsies observed in ConfirmMDx clinical utility field study. *American health & drug benefits* **7**, 129 (2014).
71 Klein, E. A. *et al.* The single-parameter, structure-based IsoPSA assay demonstrates improved diagnostic accuracy for detection of any prostate cancer and high-grade prostate cancer compared to a concentration-based assay of total prostate-specific antigen: a preliminary report. *European Urology* (2017).
72 Hessels, D. *et al.* Detection of TMPRSS2-ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. *Clinical Cancer Research* **13**, 5103-5108 (2007).
73 Tomlins, S. A. *et al.* Role of the TMPRSS2-ERG gene fusion in prostate cancer. *Neoplasia* **10**, 177IN171-188IN179 (2008).
74 Eble, J. N., Sauter, G., Epstein, J. I. & Sesterhenn, I. A. Pathology and genetics of tumours of the urinary system and male genital organs. *World Health Organization Classi* (2004).
75 Ather, M. H., Masood, N. & Siddiqui, T. Current management of advanced and metastatic renal cell carcinoma. *Urology Journal* **7**, 1-9 (2010).

76 Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E. & Ulbright, T. M. The 2016 WHO classification of tumours of the urinary system and male genital organs—part A: renal, penile, and testicular tumours. *European urology* **70**, 93-105 (2016).
77 Rodrigues, D. *et al.* Renal cell carcinoma: a critical analysis of metabolomic biomarkers emerging from current model systems. *Translational research* **180**, 1-11 (2017).
78 Lam, J. S., Leppert, J. T., Belldegrun, A. S. & Figlin, R. A. Novel approaches in the therapy of metastatic renal cell carcinoma. *World journal of urology* **23**, 202-212 (2005).
79 Najjar, Y. G. & Rini, B. I. Novel agents in renal carcinoma: a reality check. *Therapeutic advances in medical oncology* **4**, 183-194 (2012).
80 Wang, D. *et al.* Urinary volatile organic compounds as potential biomarkers for renal cell carcinoma. *Biomedical reports* **5**, 68-72 (2016).
81 Monteiro, M. *et al.* GC‐MS metabolomics‐based approach for the identification of a potential VOC‐biomarker panel in the urine of renal cell carcinoma patients. *Journal of cellular and molecular medicine* **21**, 2092-2105 (2017).
82 Morrissey, J. J. *et al.* Evaluation of urine aquaporin-1 and perilipin-2 concentrations as biomarkers to screen for renal cell carcinoma: a prospective cohort study. *JAMA oncology* **1**, 204-212 (2015).
83 Kaufman, D. S., Shipley, W. U. & Feldman, A. S. Bladder cancer. *The Lancet* **374**, 239-249 (2009).
84 Mostofi, F. Histologic typing of urinary bladder tumors. *International histological classification of tumors*, 237-256 (1973).
85 Pashos, C. L., Botteman, M. F., Laskin, B. L. & Redaelli, A. Bladder cancer: epidemiology, diagnosis, and management. *Cancer practice* **10**, 311-322 (2002).
86 Pow-Sang, J. M. & Seigne, J. D. Contemporary management of superficial bladder cancer. *Cancer Control* **7**, 335-339 (2000).
87 Cummings, K. B., Barone, J. & Ward, W. S. Diagnosis and staging of bladder cancer. *The Urologic clinics of North America* **19**, 455-465 (1992).
88 Young, M. J. & Soloway, M. S. Office evaluation and management of bladder neoplasms. *Urologic Clinics of North America* **25**, 603-611 (1998).
89 CancerNet, P. (Bethesda, Md: CancerNet PDQ, National Cancer Institute. Available at: http …, 2001).
90 Burchardt, M. *et al.* Current concepts in biomarker technology for bladder cancers. *Clinical chemistry* **46**, 595-605 (2000).
91 Zlotta, A. R. & Schulman, C. C. Biological markers in superficial bladder tumors and their prognostic significance. *Urologic Clinics of North America* **27**, 179-189 (2000).
92 Poulakis, V. *et al.* A comparison of urinary nuclear matrix protein‐22 and bladder tumour antigen tests with voided urinary cytology in detecting and following bladder cancer: the prognostic value of false‐positive results. *BJU international* **88**, 692-701 (2001).
93 Khalid, T. *et al.* A pilot study combining a GC-sensor device with a statistical model for the identification of bladder cancer from urine headspace. *PloS one* **8**, e69602 (2013).
94 Lu, N. Z. *et al.* International Union of Pharmacology. LXV. The pharmacology and classification of the nuclear receptor superfamily: glucocorticoid, mineralocorticoid, progesterone, and androgen receptors. *Pharmacological reviews* **58**, 782-797 (2006).
95 Roy, A. *et al.* Regulation of androgen action. *Vitamins & Hormones* **55**, 309-352 (1998).
96 Corbin, J. M. & Ruiz-Echevarría, M. J. One-Carbon Metabolism in Prostate Cancer: The Role of Androgen Signaling. *International journal of molecular sciences* **17**, 1208 (2016).

97 Stover, P. J. One-carbon metabolism–genome interactions in folate-associated pathologies. *The Journal of nutrition* **139**, 2402-2405 (2009).
98 Mentch, S. J. & Locasale, J. W. One‐carbon metabolism and epigenetics: understanding the specificity. *Annals of the New York Academy of Sciences* **1363**, 91-98 (2016).
99 Locasale, J. W. Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nature reviews Cancer* **13**, 572-583 (2013).
100 Hakimi, A. A. *et al.* An integrated metabolic atlas of clear cell renal cell carcinoma. *Cancer cell* **29**, 104-116 (2016).
101 Ottaviani, S. *et al.* Characterisation of the androgen regulation of glycine N-methyltransferase in prostate cancer cells. *Journal of molecular endocrinology* **51**, 301-312 (2013).
102 Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910-914 (2009).
103 Khan, A. P. *et al.* The role of sarcosine metabolism in prostate cancer progression. *Neoplasia* **15**, 491IN496-501IN413 (2013).
104 Green, T., Chen, X., Ryan, S., Asch, A. S. & Ruiz‐Echevarría, M. J. TMEFF2 and SARDH cooperate to modulate one‐carbon metabolism and invasion of prostate cancer cells. *The Prostate* **73**, 1561-1575 (2013).
105 Massie, C. E. *et al.* The androgen receptor fuels prostate cancer by regulating central metabolism and biosynthesis. *The EMBO journal* **30**, 2719-2733 (2011).
106 Barfeld, S. J., Itkonen, H. M., Urbanucci, A. & Mills, I. G. Androgen-regulated metabolism and biosynthesis in prostate cancer. *Endocrine-related cancer* **21**, T57-T66 (2014).
107 Tennakoon, J. B. *et al.* Androgens regulate prostate cancer cell growth via an AMPK-PGC-1α-mediated metabolic switch. *Oncogene* **33**, 5251-5261 (2014).
108 Tsouko, E. *et al.* Regulation of the pentose phosphate pathway by an androgen receptor–mTOR-mediated mechanism and its role in prostate cancer cell growth. *Oncogenesis* **3**, e103 (2014).
109 Costello, L., Franklin, R. & Feng, P. Mitochondrial function, zinc, and intermediary metabolism relationships in normal prostate and prostate cancer. *Mitochondrion* **5**, 143-153 (2005).
110 Costello, L., Feng, P., Milon, B., Tan, M. & Franklin, R. Role of zinc in the pathogenesis and treatment of prostate cancer: critical issues to resolve. *Prostate cancer and prostatic diseases* **7**, 111-117 (2004).
111 Franz, M.-C. *et al.* Zinc transporters in prostate cancer. *Molecular aspects of medicine* **34**, 735-741 (2013).
112 Asgari, Y., Zabihinpour, Z., Salehzadeh-Yazdi, A., Schreiber, F. & Masoudi-Nejad, A. Alterations in cancer cell metabolism: the Warburg effect and metabolic adaptation. *Genomics* **105**, 275-281 (2015).
113 Costello, L. & Franklin, R. Novel role of zinc in the regulation of prostate citrate metabolism and its implications in prostate cancer. *The Prostate* **35**, 285-296 (1998).
114 Menendez, J. A. & Lupu, R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nature Reviews Cancer* **7**, 763-777 (2007).
115 Swinnen, J. V., Brusselmans, K. & Verhoeven, G. Increased lipogenesis in cancer cells: new players, novel targets. *Current Opinion in Clinical Nutrition & Metabolic Care* **9**, 358-365 (2006).

116 Igal, R. A. Stearoyl-CoA desaturase-1: a novel key player in the mechanisms of cell proliferation, programmed cell death and transformation to cancer. *Carcinogenesis* **31**, 1509-1515 (2010).
117 Nickerson, M. L. *et al.* Improved identification of von Hippel-Lindau gene alterations in clear cell renal tumors. *Clinical cancer research* **14**, 4726-4734 (2008).
118 Du, W. *et al.* HIF drives lipid deposition and cancer in ccRCC via repression of fatty acid metabolism. *Nature communications* **8**, 1769 (2017).
119 Heinlein, C. A. & Chang, C. Androgen receptor in prostate cancer. *Endocrine reviews* **25**, 276-308 (2004).
120 Deep, G. & Schlaepfer, I. R. Aberrant lipid metabolism promotes prostate cancer: role in cell survival under hypoxia and extracellular vesicles biogenesis. *International journal of molecular sciences* **17**, 1061 (2016).
121 Griffin, J. E. Androgen resistance—the clinical and molecular spectrum. *New England Journal of Medicine* **326**, 611-618 (1992).
122 Yoshii, Y. *et al.* Fatty acid synthase is a key target in multiple essential tumor functions of prostate cancer: uptake of radiolabeled acetate as a predictor of the targeted therapy outcome. *PloS one* **8**, e64570 (2013).
123 Eidelman, E., Twum-Ampofo, J., Ansari, J. & Siddiqui, M. M. The Metabolic Phenotype of Prostate Cancer. *Frontiers in oncology* **7**, 131 (2017).
124 Tan, L. T.-H. *et al.* Targeting membrane lipid a potential cancer cure? *Frontiers in pharmacology* **8** (2017).
125 Ekwueme, D. U., Stroud, L. A. & Chen, Y. Peer Reviewed: Cost Analysis of Screening for, Diagnosing, and Staging Prostate Cancer Based on a Systematic Review of Published Studies. *Preventing chronic disease* **4** (2007).
126 Aubry, W. *et al.* Budget impact model: epigenetic assay can help avoid unnecessary repeated prostate biopsies and reduce healthcare spending. *American health & drug benefits* **6**, 15 (2013).
127 Van Neste, L. *et al.* The epigenetic promise for prostate cancer diagnosis. *The Prostate* **72**, 1248-1261 (2012).
128 HB, S. *What's the downside to a biopsy? Prostate Knowledge, Harvard Medical School publications.* https://www.harvardprostateknowledge.org/whats-the-downside-to-a-biopsy. *Accessed May 23, 2018.*
129 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *cell* **144**, 646-674 (2011).
130 Silva, C. L., Passos, M. & Câmara, J. S. Solid phase microextraction, mass spectrometry and metabolomic approaches for detection of potential urinary cancer biomarkers—a powerful strategy for breast cancer diagnosis. *Talanta* **89**, 360-368 (2012).
131 Willis, C. M. *et al.* Olfactory detection of human bladder cancer by dogs: proof of principle study. *Bmj* **329**, 712 (2004).
132 Taverna, G. *et al.* Olfactory system of highly trained dogs detects prostate cancer in urine samples. *The Journal of urology* **193**, 1382-1387 (2015).
133 Sonoda, H. *et al.* Colorectal cancer screening with odour material by canine scent detection. *Gut* **60**, 814-819 (2011).
134 Matsumura, K. *et al.* Urinary volatile compounds as biomarkers for lung cancer: a proof of principle study using odor signatures in mouse models of lung cancer. *PloS one* **5**, e8819 (2010).

135 Amann, A. *et al.* The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. *Journal of breath research* **8**, 034001 (2014).
136 Fan, J. & Lv, J. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 849-911 (2008).
137 Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288 (1996).
138 Kleinbaum, D. a. K., M. . *Logistic Regression, A Self-Learning Text. 3rd Edition.*, (2010).
139 Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27-38 (1993).
140 López-Ratón, M., Rodríguez-Álvarez, M. X., Cadarso-Suárez, C. & Gude-Sampedro, F. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software* **61**, 1-36 (2014).
141 Team, R. C. *R: A language and environment for statistical computing*, <https://www.R-project.org/> (2017).
142 Chistiakov, D. A., Myasoedova, V. A., Grechko, A. V., Melnichenko, A. A. & Orekhov, A. N. in *Seminars in cancer biology*. (Elsevier).
143 Stewart, G. D. *et al.* Clinical utility of an epigenetic assay to detect occult prostate cancer in histopathologically negative biopsies: results of the MATLOC study. *The Journal of urology* **189**, 1110-1116 (2013).
144 Trock, B. J. in *Urologic Oncology: Seminars and Original Investigations*. 572-581 (Elsevier).
145 Crowe, F. L. *et al.* 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–. *The American journal of clinical nutrition* **88**, 1353-1363 (2008).
146 Epstein, M. M. *et al.* Dietary fatty acid intake and prostate cancer survival in Örebro County, Sweden. *American journal of epidemiology* **176**, 240-252 (2012).
147 Kim, S. *et al.* Myristoylation of Src kinase mediates Src induced and high fat diet accelerated prostate tumor progression in mice. *Journal of Biological Chemistry*, jbc. M117. 798827 (2017).
148 Nadler, M. J., Harrison, M. L., Ashendel, C. L., Cassady, J. M. & Geahlen, R. L. Treatment of T cells with 2-hydroxymyristic acid inhibits the myristoylation and alters the stability of p56lck. *Biochemistry* **32**, 9250-9255 (1993).
149 Noguchi, M. *et al.* Induction of cellular and humoral immune responses to tumor cells and peptides in HLA‐A24 positive hormone‐refractory prostate cancer patients by peptide vaccination. *The Prostate* **57**, 80-92 (2003).
150 Harada, M., Noguchi, M. & Itoh, K. Target molecules in specific immunotherapy against prostate cancer. *International journal of clinical oncology* **8**, 193-199 (2003).
151 National Center for Biotechnology Information. PubChem Compound Database; CID=73219, h. p. n. n. n. g. c.
152 Zhong, C., Yang, S., Huang, J., Cohen, M. B. & Roy-Burman, P. Aberration in the expression of the retinoid receptor, RXRα, in prostate cancer. *Cancer biology & therapy* **2**, 179-184 (2003).
153 Gann, P. H., Hennekens, C. H., Ma, J., Longcope, C. & Stampfer, M. J. Prospective study of sex hormone levels and risk of prostate cancer. *JNCI: Journal of the National Cancer Institute* **88**, 1118-1126 (1996).

154 Cooperberg, M. R. Prostate cancer risk assessment: choosing the sharpest tool in the shed. *Cancer: Interdisciplinary International Journal of the American Cancer Society* **113**, 3062-3066 (2008).
155 Scher, H. I. & Heller, G. Clinical states in prostate cancer: toward a dynamic model of disease progression. *Urology* **55**, 323-327 (2000).
156 Chang, A. J., Autio, K. A., Roach III, M. & Scher, H. I. High-risk prostate cancer—classification and therapy. *Nature reviews Clinical oncology* **11**, 308 (2014).
157 Fleming, C. *et al.* A decision analysis of alternative treatment strategies for clinically localized prostate cancer. *Jama* **269**, 2650-2658 (1993).
158 Cooperberg, M. R., Broering, J. M. & Carroll, P. R. Time trends and local variation in primary treatment of localized prostate cancer. *Journal of Clinical Oncology* **28**, 1117 (2010).
159 Wilt, T. J. *et al.* Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Annals of internal medicine* **148**, 435-448 (2008).
160 Thompson, I. M. *et al.* Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. *Journal of the National Cancer Institute* **98**, 529-534 (2006).
161 Tomlins, S. A. *et al.* Urine TMPRSS2: ERG fusion transcript stratifies prostate cancer risk in men with elevated serum PSA. *Science translational medicine* **3**, 94ra72-94ra72 (2011).
162 Rastinehad, A. R. *et al.* D'Amico risk stratification correlates with degree of suspicion of prostate cancer on multiparametric magnetic resonance imaging. *The Journal of urology* **185**, 815-820 (2011).
163 Dufort, I., Soucy, P. & Lacoste, L. Comparative biosynthetic pathway of androstenol and androgens. *The Journal of steroid biochemistry and molecular biology* **77**, 223-227 (2001).
164 Devgan, S. A. *et al.* Genetic variation of 3β‐hydroxysteroid dehydrogenase type II in three racial/ethnic groups: Implications for prostate cancer risk. *The Prostate* **33**, 9-12 (1997).
165 Thomas, L. N. *et al.* Differential alterations in 5α‐reductase type 1 and type 2 levels during development and progression of prostate cancer. *The Prostate* **63**, 231-239 (2005).
166 Forman, B. M. *et al.* Androstane metabolites bind to and deactivate the nuclear receptor CAR-β. *Nature* **395**, 612 (1998).
167 Moore, L. B. *et al.* Orphan nuclear receptors constitutive androstane receptor and pregnane X receptor share xenobiotic and steroid ligands. *Journal of Biological Chemistry* **275**, 15122-15127 (2000).
168 Fujimura, T. *et al.* Clinical significance of steroid and xenobiotic receptor and its targeted gene CYP3A4 in human prostate cancer. *Cancer science* **103**, 176-180 (2012).
169 Hernandez, D. J., Nielsen, M. E., Han, M. & Partin, A. W. Contemporary evaluation of the D'amico risk classification of prostate cancer. *Urology* **70**, 931-935 (2007).
170 Kattan, M. W., Eastham, J. A., Stapleton, A. M., Wheeler, T. M. & Scardino, P. T. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *JNCI: Journal of the National Cancer Institute* **90**, 766-771 (1998).
171 Cooperberg, M. R. *et al.* The University of California, San Francisco Cancer of the Prostate Risk Assessment score: a straightforward and reliable preoperative predictor of disease recurrence after radical prostatectomy. *The Journal of urology* **173**, 1938-1942 (2005).
172 Rini, B. I., Campbell, S. C. & Escudier, B. Renal cell carcinoma. *The Lancet* **373**, 1119-1132 (2009).

173 Weiss, R. H. & Kim, K. Metabolomics in the study of kidney diseases. *Nature Reviews Nephrology* **8**, 22 (2012).
174 McMahon, G. M. & Waikar, S. S. Biomarkers in nephrology: core curriculum 2013. *American Journal of Kidney Diseases* **62**, 165-178 (2013).
175 Andreoli, R., Manini, P., Corradi, M., Mutti, A. & Niessen, W. M. Determination of patterns of biologically relevant aldehydes in exhaled breath condensate of healthy subjects by liquid chromatography/atmospheric chemical ionization tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **17**, 637-645 (2003).
176 Liu, H. *et al.* Nitro-oleic acid protects the mouse kidney from ischemia and reperfusion injury. *American Journal of Physiology-Renal Physiology* **295**, F942-F949 (2008).
177 Nemer, M. J. & Elwyn, D. The conversion of serine to ethanolamine and its derivatives in the rat. *Journal of Biological Chemistry* **235**, 2070-2074 (1960).
178 Taylor, J. M., Ankerst, D. P. & Andridge, R. R. Validation of biomarker-based risk prediction models. *Clinical Cancer Research* **14**, 5977-5983 (2008).
179 Pepe, M. S. *et al.* Phases of biomarker development for early detection of cancer. *Journal of the National Cancer Institute* **93**, 1054-1061 (2001).

Vita

Qin Gao earned her Bachelor of Science degree in pharmacy from China Pharmaceutical University in 2010. In 2013, she received her Master of Science degree in pharmacology from China Pharmaceutical University. Then, she joined UTEP's doctoral program in chemistry and biochemistry in 2014.

She was the recipient of Wiemer Family Student Endowment for Excellence Scholarship, Dodson Research Grant, and Dr. Keelung Hong Chemistry Graduate Research Fellowship.

She has presented her research at several conferences including 2016 Graduate Research Expo in UTEP, 2017 Pittcon Conference and Expo, 2017 BBRC (Border Biomedical Research Center) Symposium, 2017 annual meeting of SBUR (Society for Basic Urologic Research), 2018 Chemistry Day Symposium of UTEP, and 2018 66th ASMS (American Society for Mass Spectrometry) Conference. Her work has been published on Genitourinary Cancer.

While pursuing her degree, she worked as a teaching assistant and research associate for the Department of Chemistry and Biochemistry. After PhD graduation, she will continue to work on metabolomics study for disease diagnosis and prognosis and push forward the commercialization of her findings in research.

Her dissertation, "Application of Urinary Metabolites in Cancer Detection", was supervised by Dr. Wen-Yee Lee.

This dissertation was typed by Qin Gao (Email: qgao@miners.utep.edu, qingaocn@live.com).