Study of plasma proteome pattern using matrix-assisted laser-desorption ionization time of flight mass spectrometry in a cohort of Egyptian prostate cancer patients

Moataz S.M. Sadaka[a], Ola A. Sharaki[a], Ahmed A. ElAbbady[b], Eman S. Nassar[a], Mohamed M.M. Rizk[a]

Departments of [a]Clinical & Chemical Pathology, [b]Urology, Faculty of Medicine, University of Alexandria, Alexandria, Egypt

Correspondence to
Moataz S.M. Sadaka, MD,
Department of Clinical & Chemical Pathology, Faculty of Medicine, University of Alexandria', 34, Bahaa Eldeen Elghatwary Street, Sidi Gaber, 704 Alexandria, Egypt
Postal Code: 21311
Tel: +20 101 001 7374/20 122 140 8224;
E-mail: m_mohamed121@alexmed.edu.eg

Received 10 February 2020
Accepted 13 April 2020
Published 04 August 2022

The Egyptian Journal of Laboratory Medicine 2020, 3:84–93

Purpose: The goal of the study was to detect the existence of an abnormal proteomic pattern in prostate cancerpatients by comparing them to control subjects from a matching age group. **Method:** The study was conducted on50 newly diagnosed prostate cancer (PCa) patients and 50 normal individuals who were proven clinically and bylaboratory investigations to be healthy. Plasma samples were processed using MB-WCX magnetic beads andsamples were tested using a mass spectrometer. Mass spectra was acquired using Bruker's FLEX-analysisprogram and later the peaks were analysed using the ClinProTools analysis. Three statistical models (Geneticalgorithm GA, Supervised Neural Network SNN, Quickclassifier QC) were generated to detect the peaks. **Results:** The results showed 26 peaks were found to be significantly expressed between the cases and thecontrols with PWKW <0.05. Out of these, six peaks were over expressed in cases, while 20 peaks were under-expressed. The GA model generated the best peak combination, showing five peaks with the m/z ratios 2485.97,1061.24, 3295.1, 4612.54 and 2817.28. This model achieved a sensitivity of 87.5% and a specificity of 92.9%during external validation. **Conclusion:** It can be concluded that proteomic profiling can be an effective method for the discovery of newblood-based tumour markers for PCa patients. Moreover, MALDI-TOF proteomic profiling represents a newfrontier for screening and early diagnosis of prostate cancer in Egypt. The analytical performance of multiplexedtumour profiles exceeds that of the single traditional tumour markers.

Keywords:
biomarkers, MALDI-TOF, mass spectrometry, prostate cancer, proteomics

Egyptian Journal of Laboratory Medicine 3:84–93

Introduction

Prostate cancer (PCa) is the second most common cancer worldwide and believed to be the sixth highest cause of mortality among men [1]. Incidence elevates as the population of males above 50 years of age increases. However, this is attributed to the improvement of diagnostic modalities such as prostatic biopsy and prostate-specific antigen (PSA) and increased screening rather than a real increase in incidence [2].

It has been reported that the incidences of PCa in the Middle East, including Egypt, are lower than the western world. It was noted in the last decade by urologists, however, that there has been an increase in diagnosed cases of PCa. This might indicate an increase in awareness and an improvement of diagnostic methods or increased incidence [3]. In Egypt in the year 2008, ~1661 men were diagnosed with PCa, with a rate of 6.6 diagnoses per 100 000, and 1283 men were expected to die from this disease at a rate of 5.1 deaths per 100 000 men [2].

The primary risk factors are age, family history, and obesity. It is uncommon for PCa to occur in a patient before 50 years of age. Nonetheless, 30% of men between 55 and 64 years of age are reported to have PCa [4].

A number of associations have been shown between some specific gene polymorphisms and the risk of PCa development, such as prostate cancer gene 3 (PCa3), alpha–methyl–acyl CoA racemase (AMACR), and TMPRSS2:ERG gene fusion [5].

Lower blood levels of vitamin D are also thought to increase the risk of PCa, as do high levels of dietary fat, smoking, alcohol consumption, and the contraction of sexually transmitted diseases [6].

More than 95% of primary PCa are adenocarcinomas and are often found to be multifocal and heterogeneous in patterns of differentiation [7]. PCa rarely causes symptoms at an early stage. The presence of symptoms suggests locally advanced or metastatic disease [8].

Screening helps in PCa detection in patients without any evident symptoms and has resulted in a decrease in PCa mortality and incidence increase. The current Food and Drug Administration guidelines for PCa diagnosis depend on PSA detection in blood together with digital rectal examination for men over 50 years of age. The recommended age at which to begin screening for men at average risk who have a minimum life expectancy of 10 years is 50 years of age. In African–American men and men who have one first-degree relative diagnosed with PCa, the recommended age at which to begin screening is 40–45 years of age. Men who have several first-degree relatives diagnosed with PCa at an early age, on the other hand, are best screened from the age of 40 [9].

It is important to note that PSA is prostate specific but not cancer specific. The presence of prostate diseases such as PCa, benign prostatic hyperplasia (BPH), and prostatitis is the most important factor affecting serum PSA levels [10]. However, not all men with prostate diseases have increased PSA levels, and PSA elevations are not cancer specific [11].

One of the occasional problems with relying on PSA as a diagnostic indicator is its low sensitivity. For example, in cases where PSA levels are between 4.0 and 10.0 ng/ml, the positive predictive value is ~25%, which is decreased compared with patients with PSA level greater than 10 ng/ml [12]. New modalities have been developed using PSA as a consequence, including free PSA, PSA velocity, and PSA volume [13]. Nonetheless, prostate biopsy remains the only definitive diagnostic modality and is held to be the gold standard in PCa diagnosis [14].

Given the low sensitivity and specificity of the currently used tumor marker PSA, however, there is a pressing need for the development of novel tumor markers that would be helpful in improving cancer diagnosis, prognosis, and treatment [15].

The term 'proteome' was first formulated in 1994 by Mark Wilkins to describe the complete set of proteins that ultimately results from genome transcription in a given cell, tissue, or organism [16]. Proteomics is the study of proteome, which is the protein complement of the genome. A cell will have only one genome but can have many proteomes, depending on which genes are expressed and the level of that expression at a particular time [17].

Proteomics is used nowadays to diagnose diseases using specific protein-biomarker discovery. Several proteomic methods help test for proteins produced during a particular disease, which aids in earlier disease diagnosis. Techniques include Western blot, immune-histochemical staining, enzyme-linked immunosorbent assay, or mass spectrometry [18].

The diagnostic capability of matrix-assisted laser-desorption ionization time of flight (MALDI/TOF) is exemplified by its use in assisting in the diagnosis of pancreatic cancer, which is one of the most difficult cancers to diagnose and with high mortality rates [19].

Protein-separation techniques are based on different physical properties of a protein. In the last decade, multiple studies have been carried out using magnetic beads as a method for offline serum peptide or protein capture in a procedure referred to as solid-phase extraction [4]. Magnetic beads are specifically designed to manually or automatically fractionate proteins or peptides from complex biological samples in an efficient manner. The most applied beads in studies are weak cation-exchange beads (WCX), reverse-phase C18 beads (RPC18), and C8 beads. WCX beads separate proteins based on charge, whereas RPC18 beads separate proteins and peptides via strong hydrophobic interaction [20].

Mass spectrometer used in this study is the MALDI–TOF. The ions that are produced are then transferred through a vacuum tube to the mass analyzer where they are separated according to their mass-to-charge ratio (m/z). Each ion usually has a single charge ($z = 1$). Thus, the m/z ratio is equal to the mass, which means that the mass is the variable that determines the time of flight and that impacts the separation. All mass analyzers measure physical entity as the m/z value of the ions. The intensity at different m/z values is the output recorded at the detector. The result is visualized as a m/z versus intensity plot called the mass spectrum [21].

The MALDI technique is an effective means of ionizing peptides and proteins. Moreover, it has been reported by some studies that the MALDI–TOF profile is better than that of the surface-enhanced laser ionization SELDI-TOF in terms of a larger number of mass-to-charge peaks.

Patients and methods

The study was conducted on 50 newly diagnosed PCa patients who were admitted to Alexandria Main University Hospital at the Urology Department between November 2015 and April 2016. The controls were 50 normal males of the same-age group who were proven clinically and by laboratory testing to be healthy. Freshly diagnosed PCa patients were chosen,

so patients who already underwent treatment by way of total surgical resection, chemotherapy, or radiation, were excluded from the study. Informed consent was taken from each participant. Each group of patients and controls was split at random into a training set of 26 patients and a validation set of 24 patients. The study was approved by the Research Ethics Committee of Alexandria University Hospital.

All participants in the study were subjected to history taking and routine laboratory investigations, which included alanine aminotransferase, aspartate aminotransferase, urea, creatinine, fasting blood sugar and complete blood-count tests, and PSA [22]. All PCa patients were confirmed by histopathological examination of prostatic tissue biopsy. Blood samples of around 5 ml were drawn from each candidate on a K_2 EDTA tube and then transferred in an ice bag to the laboratory. Each sample was centrifuged in a cold centrifuge at 4°C for 15 min at 1800 g. The plasma was then separated and pipetted into five DNA low-bind Eppendorf tubes, which were subsequently numbered and stored in a box at −80°C.

Before choosing which type of beads to use, a pilot study composed of plasma samples of 10 PCa patients and 10 controls was performed using two types of magnetic beads – MB-WCX and MB-C8 beads – in which WCX showed better results in comparison with C8 with respect to peak acquisition and abundance, it also yielded better validation results, higher overall recognition capability, and cross-validation using WCX-MBs compared with C8-MBs as shown in Table 1.

Plasma samples (5 μl) were processed using MB-WCX kits from BRUKER Daltonics, and steps for protein purification were followed according to the manufacturer's manual. In all, 1 μl of the resulting elute was spotted on a polished steel-target plate, then 1 μl of matrix consisting of α-cyano-4-hydroxycinnamic acid (HCCA 3 mg/ml) in 50% acetonitrile 2% TFA was added to the dried-sample elute [23].

To evaluate reproducibility, each plasma sample was spotted at four positions. The air-dried targets were readied to be analyzed by the ultrafleXtreme mass spectrometer using the flexControl analysis program. The program settings were adjusted so that the detection limit was set to 800–20 000 Da, and spectral acquisition was performed in the positive linear mode. Only peaks with a signal/noise ratio above three were chosen from the generated spectra and then mass spectrometry was operated in information-dependant analysis mode.

Most importantly, calibration was done using the ClinProTool Calibrant Standards (CPA) before testing the samples.

For data processing, the peaks acquired from the flexControl analysis program were analyzed using the ClinProTools V 3.0 program, which was able to perform spectra pretreatment, peak picking, and peak-calculation operations, and was used to recognize peptide patterns.

Comparisons between PCa patients and healthy controls were performed using the Wilcoxon test; statistical significance was assumed when the *P* value was less than 0.05. All spectra were loaded for peak statistic calculation, model generation, and classification. Three different statistical models were used to determine the most optimal peaks, being the genetic algorithm (GA), supervised neural network (SNN), and quick classifier (QC) models.

Results

We analyzed the serum peptidome fingerprints of all 50 patients with PCa as well as the 50 healthy controls. Further, we evaluated changes at the peptidome level in the plasma samples of 50 PCa patients compared with 50 healthy controls in the training set. By analyzing the spectra (screened from two groups in the training set) using ClinProTools software V3.0, we were able to identify proteomic patterns that distinguished between PCa patients and the healthy controls.

GA settings have been set to use the first 100 peaks (according to PWKW) for model generation, 69 peaks have been found to be true peaks as shown in Table 2. After checking the peak average area/intensity for each class in Table 3, 26 peaks were found to be significantly expressed between the patients and

Table 1 Results of pilot study between WCX-MB and C8-MB

Name algorithm	Validation			Name	Algorithm	Validation		
	XVal	X1×2	Recognition capability			XVal	X1×2	Recognition capability
WCX-MB GA	95%	100% 90%	100%	C8-MB	GA	80%	70% 90%	100%
SNN	85%	90% 80%	100%		SNN	80%	80% 80%	95%
QC	85%	80% 90%	100%		QC	80%	90% 70%	90%

GA, genetic algorithm; QC, quick classifier; SNN, supervised neural network; WCX, weak cation-exchange bead; Xval, overall cross-validation; X1, cross-validation 1; X2, cross-validation 2.

Table 2 The peak statistics table of the peaks that have been used to differentiate prostate cancer patients from controls

ClinProTools version				3.0 build 22		
Number of peaks				69		
Sort mode				*P* WKW		
S	Index	Mass	DAve	PTTA	PWKW	PAD
X	11	2485.97	30.9	<0.000001	<0.000001	0.0000344
X	6	1553.97	2.61	0.00173	0.000219	0.976
X	12	2660.79	6.21	0.0049	0.00104	<0.000001
X	27	4419.49	3.64	0.0471	0.00187	<0.000001
X	28	4435.4	2.95	0.0363	0.00344	<0.000001
X	8	2288.31	2.91	0.00306	0.00896	0.159
X	53	9135.46	0.21	0.0049	0.0101	0.00112
X	48	7566.2	1.71	0.0162	0.0108	<0.000001
X	29	4457.48	1.47	0.0249	0.0133	<0.000001
X	17	2878.94	20.31	0.0125	0.0149	0.296
X	31	4575.05	17.73	0.00632	0.0149	0.00405
X	23	3279.43	4.81	0.0049	0.016	0.00000116
X	54	9178.04	0.21	0.00415	0.016	0.0000401
X	59	9948.62	0.11	0.00632	0.016	0.255
X	64	11099.79	0.18	0.0597	0.0202	0.00000182
X	69	16613.29	0.25	0.0681	0.0213	<0.000001
X	24	3295.1	1.32	0.0162	0.0253	0.000312
X	15	2831.59	5.06	0.0126	0.027	0.0294
X	68	15127.93	0.41	0.0205	0.027	<0.000001
X	65	11730.75	0.14	0.0993	0.027	<0.000001
X	63	11080.01	0.21	0.056	0.029	<0.000001
X	47	6960.37	0.19	0.0993	0.029	<0.000001
X	25	4093.3	0.58	0.0222	0.0427	0.0287
X	52	8565.91	1.26	0.0597	0.0459	<0.000001
X	51	8528.63	0.25	0.0294	0.0459	<0.000001
X	9	2357.77	7.25	0.0597	0.0489	0.0443
X	3	1061.24	3.44	0.0232	0.0521	0.00786
X	20	2939.1	3.37	0.0316	0.0584	0.0232
X	1	925.11	5.96	0.0249	0.0592	0.00000164
X	30	4535.19	0.66	0.026	0.0621	0.00137
X	66	13293.81	0.22	0.111	0.0621	<0.000001
X	67	13315.56	0.2	0.154	0.0621	<0.000001
X	49	7648.68	0.25	0.0162	0.0694	<0.000001
X	32	4590.4	4.21	0.0363	0.0706	0.00634
X	44	6632.5	1.1	0.056	0.0719	<0.000001
X	26	4407.3	0.89	0.0968	0.0732	<0.000001
X	39	5669.39	0.33	0.0595	0.0817	0.636
S	Index	Mass	DAve	PTTA	PWKW	PAD
X	38	5002.59	1.02	0.0268	0.095	<0.000001
X	43	6476.72	0.24	0.0405	0.122	0.0165
X	60	10106.33	0.08	0.056	0.122	<0.000001
X	14	2817.28	2.41	0.0993	0.129	0.201
X	58	9423.01	0.13	0.0595	0.137	<0.000001
X	56	9324.58	0.07	0.295	0.183	0.0163
X	21	3241.69	9.89	0.127	0.215	0.000703
X	18	2901.1	9.12	0.124	0.299	0.0563
X	7	1981.62	5.31	0.0425	0.323	<0.000001
X	2	947.58	4.64	0.154	0.337	0.0048
X	33	4612.54	2.74	0.243	0.351	<0.000001
X	41	5904.41	0.2	0.127	0.378	0.000946
X	42	6433.5	0.24	0.279	0.409	<0.000001
X	46	6664.88	0.02	0.969	0.409	0.00182
X	45	6648.41	0.13	0.56	0.425	0.00185
X	61	10445.28	0.15	0.0993	0.429	<0.000001
X	40	5866.51	0.03	0.969	0.497	<0.000001
X	13	2802.26	1.26	0.219	0.523	0.00642
X	62	10836.34	0.04	0.623	0.523	<0.000001
X	36	4964.76	0.03	0.988	0.523	<0.000001
X	37	4977.95	0.33	0.522	0.582	<0.000001
X	16	2863.3	7.57	0.771	0.644	0.277
X	55	9288.43	0.23	0.467	0.663	0.0000506
X	19	2916.91	0.45	0.822	0.712	0.121
X	10	2423.28	1.67	0.462	0.808	0.00000302
X	4	1098.85	0.39	0.969	0.811	<0.000001
X	5	1136.3	1.04	0.706	0.892	0.000007
X	35	4647.93	0.14	0.706	0.894	0.041
X	50	7766.16	0.22	0.378	0.927	<0.000001
X	34	4628.31	0.47	0.513	0.929	<0.000001
X	57	9362.56	0	0.988	0.931	0.36
X	22	3257.17	0.08	0.988	0.962	0.0414

Contd...

Table 2 Contd...

S, inclusion/exclusion state of the peak; X means peak is included for model generation while - means a peak is not included for model generation. Here, all peaks have been used. Index: peak index. Mass: *m/z* ratio. Dave, difference between the maximal and the minimal average peak area/intensity of all classes. PTTA *P* value of *t* test (two classes) or analysis of variance test (>2 classes), range 0-1; 0: good, 1: bad. Preferable for normal distributed data. PWKW *P* value of Wilcoxon test (two classes) or Kruskal-Wallis test (>2 classes), range 0-1; 0: good, 1: bad. Preferable for not normal distributed data. PAD *P* value of Anderson-Darling test gives information about normal distribution; range 0-1; 0: not normal distributed, 1: normal distributed.

the controls with PWKW less than 0.05 as shown in Table 3. Out of these, six peaks were upexpressed in patients, while 20 peaks were downexpressed in patients in relation to controls. The PWKW illustrates the results from the univariate point of view, these peaks can discriminate between PCa patients and healthy controls. However, the GA model has a multivariate-detection capability, where the perfect number of peaks that constitutes a ClinProt Model was set in GA to 5. The model chooses the best combination of five peaks that discriminates between the two classes.

Using GA to build the differentiating model, five peaks were found, which represented the proteomic profile that differentiated between both groups. These five peaks had the *m/z* ratios 2485.97, 1061.24, 3295.1, 4612.54, and 2817.28, whereby four were underexpressed in patients, while one (peak 24 = 3295.1) was overexpressed. In addition to its multivariate discriminatory power as a profile, three peaks were found through univariate analysis to be statistically significant and two were found to be insignificant between both groups. Their *P* values were as follows: peak 11 (*m/z* 2485.97) less than 0.000001, *P* value of peak 3 (*m/z* 1061.24)=0.05, and *P* value of

peak 24 (*m/z* 3295.1)=0.0253, contrary to the *P* value of peak 33 (*m/z* 4612.54)=0.351 and *P* value of peak 14 (*m/z* 2817.28)=0.129.

The GA model demonstrated a higher recognition capability (100%) and higher cross-validation (Xval = 91.5%) than both the SNN model (RC = 100%, XVal = 84%) and QC model (RC = 98.2%, XVal = 86.4%). Moreover, the external validation of the GA model was higher than the other two models, showing sensitivity and specificity of 87.5 and 92.9%, respectively, as opposed to 100 and 64.3% in the SNN model and 87.5 and 92.9% in the QC model.

The results of our study were plotted as whole spectral view (Fig. 1) showing the peaks from the two classes, also peak expressions (Fig. 2,3) showing the difference in expression of peaks 11 and 24 between patients and healthy participants.

Another plot was the box and whiskers. The top-end and bottom-end marks of the plot (whiskers) indicate the maximum and minimum peak intensity within a given class. The box indicates the 25% quartile (bottom) and the 75% quartile (top) and the horizontal intersection denotes the median. About 50% of the values fall into this interquartile range and the whiskers give you an impression of how much the remaining 50% of the values spread. The plot allows assessment of the quality of the peaks in a model. A peak where the box and whiskers of the individual classes are well separated with only minimal overlap of the whiskers is better than the overlapping ones. As shown in Fig. 4, scaling for the box-and-whiskers command is unique and independent of the peak-intensity scale.

The 2D peak statistics (Fig. 5) that uses two peaks to separate between the two studied classes uses the first two peaks (default) or manually choosing two peaks to be used. In this study, we used the first two peaks and the five peaks of the model against each other to see which two peaks can best separate PCa patients from controls as shown in Fig. 5, where peaks 6 and 11 were used.

The 2D peak statistics (Fig. 5) that uses two peaks to separate between the two studied classes uses the first two peaks (default) or manually choosing two peaks to be used. In this study, we used the first two peaks and the five peaks of the model against each other to see which two peaks can best separate PCa patients from controls as shown in Fig. 5, where peaks 6 and 11 were used.

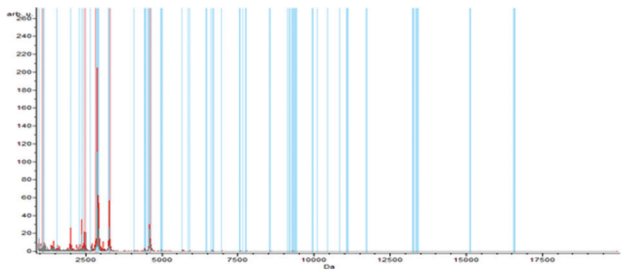
Area under the curve (AUC) of the receiver-operating characteristic curve is a combined measure of sensitivity and specificity, and it represents an important tool for the analysis of the performance of diagnostic tests [24]. For the peaks generated in our profile, we achieved an AUC = 0.725 of the peak with *m/z* 3295.1 Da as shown in Fig. 6, an AUC = 0.690994 of the peak with *m/z* 1061.24 Da, and an AUC = 1.0 of peak 11 with *m/z* 2485.97 Da.

Part of the present study attempted to detect the abnormality in the plasma of patients in the gray zone (PSA = 4–10 ng/ml) – the stage in which all patients, whether they are ultimately found to have cancer, are more prone to undergo prostatic biopsy for diagnosis – by testing 10 gray-zone patients against 25 controls. GA was used to create a model or profile to detect significant resultant peaks from the analysis. The following peaks with *m/z* ratios were identified: peak 11 = 2485.88, peak 3 = 1061.19, peak 6 = 1553.91, peak 22 = 3295.1, and peak 19 = 3241.58. The whole multivariate model could distinguish PCa patients from gray-zone patients with a recognition capability of 100% and overall cross-validation of 99.38%. Only peaks 11 and 3 were significant in the univariate analysis.

Discussion

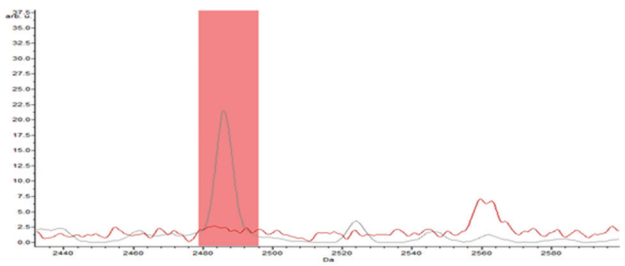
Mass spectrometry has increasingly been used as an approach to select new marker candidates for

Figure 1



The whole spectral view in ClinProTools. The figure represents class I (patients) in red against class II (control) in gray.

Figure 2



Peak 11 expression between patients and controls. The red line demarcates the peak as an integration region with start mass of 2478.63 and end mass of 2496.

Table 3 Peak statistics table showing the 26 significant peaks

ClinProTools version				3.0 build 22								
Number of peaks				26								
Sort mode				*P* WKW								
S	Index	Mass	DAve	PTTA	PWKW	PAD	Ave1	Ave2	StdDev1	StdDev2	CV1	CV2
X	11	2485.97	30.9	<0.000001	<0.000001	0.0000344	7.36	38.27	3.63	15.7	49.26	41.02
X	6	1553.97	2.61	0.00173	0.000219	0.976	5.7	8.3	2.22	1.83	38.95	22.04
X	12	2660.79	6.21	0.0049	0.00104	<0.000001	10.55	4.34	6.9	3.4	65.38	78.26
X	27	4419.49	3.64	0.0471	0.00187	<0.000001	5.87	2.23	6.89	0.75	117.26	33.51
X	28	4435.4	2.95	0.0363	0.00344	<0.000001	4.59	1.63	5.21	0.72	113.63	43.98
X	8	2288.31	2.91	0.00306	0.00896	0.159	4.6	7.51	1.95	3	42.48	39.94
X	53	9135.46	0.21	0.0049	0.0101	0.00112	0.34	0.55	0.13	0.25	37.92	45.27
X	48	7566.2	1.71	0.0162	0.0108	<0.000001	0.38	2.1	0.14	2.76	36.77	131.73
X	29	4457.48	1.47	0.0249	0.0133	<0.000001	2.7	1.22	2.35	0.43	86.99	35.46
X	17	2878.94	20.31	0.0125	0.0149	0.296	38.9	59.21	20.4	23.54	52.43	39.76
X	31	4575.05	17.73	0.00632	0.0149	0.00405	16.68	34.41	11.25	22.58	67.46	65.64
X	23	3279.43	4.81	0.0049	0.016	0.00000116	9.27	4.46	5.53	2.07	59.62	46.45
X	54	9178.04	0.21	0.00415	0.016	0.0000401	0.37	0.58	0.11	0.25	29.72	43.23
X	59	9948.62	0.11	0.00632	0.016	0.255	0.29	0.4	0.11	0.1	40.05	26.28
X	64	11099.79	0.18	0.0597	0.0202	0.00000182	0.29	0.47	0.24	0.34	80.76	71.58
X	69	16613.29	0.25	0.0681	0.0213	<0.000001	0.15	0.4	0.18	0.57	121.06	142.99
X	24	3295.1	1.32	0.0162	0.0253	0.000312	4.32	3.01	1.71	1.17	39.6	38.98
X	15	2831.59	5.06	0.0126	0.027	0.0294	7.74	12.8	4.02	6.84	52	53.47
X	68	15127.93	0.41	0.0205	0.027	<0.000001	0.06	0.48	0.03	0.71	49.9	147.39
X	65	11730.75	0.14	0.0993	0.027	<0.000001	0.29	0.43	0.26	0.26	88.51	61.06
X	63	11080.01	0.21	0.056	0.029	<0.000001	0.29	0.5	0.25	0.38	85.17	76.87
X	47	6960.37	0.19	0.0993	0.029	<0.000001	0.79	0.98	0.28	0.41	35.4	41.52
X	25	4093.3	0.58	0.0222	0.0427	0.0287	1.9	2.47	0.55	0.85	29.19	34.24
X	52	8565.91	1.26	0.0597	0.0459	<0.000001	0.52	1.78	0.27	2.92	52.98	164.26
X	51	8528.63	0.25	0.0294	0.0459	<0.000001	0.41	0.66	0.16	0.45	38.65	67.93
X	9	2357.77	7.25	0.0597	0.0489	0.0443	16.55	23.8	11.31	11.9	68.33	49.99
X	3	1061.24	3.44	0.0232	0.0521	0.00786	8.21	11.65	2.97	5.33	36.22	45.77

S: inclusion/exclusion state of the peak; X means peak is included for model generation while - means a peak is not included for model generation. Here, all peaks have been used. Index: peak index. Mass: *m/z* ratio. Dave, difference between the maximal and the minimal average peak area/intensity of all classes. PTTA *P* value of *t* test (two classes) or analysis of variance test (>2 classes), range 0-1; 0: good, 1: bad. Preferable for normal distributed data. PWKW *P* value of Wilcoxon test (two classes) or Kruskal-Wallis test (>2 classes), range 0-1; 0: good, 1: bad. Preferable for not normal distributed data. PAD *P* value of Anderson-Darling test gives information about normal distribution; range 0-1; 0: not normal distributed, 1: normal distributed. Ave*N*: peak average area/intensity for class N. Peak intensity was chosen as the sort mode in the average peak list calculation settings. So Ave1 is the average intensity for class I (cases) and Ave2 is the average intensity for class II (control). stdDev*N*, SD of the peak average intensity for class N. CVN, coefficient of variation in % of class N.

early disease diagnosis, the monitoring of disease progression, and to assess the therapeutic effects of drugs. Technological breakthroughs allowing profiling of the entire proteome are promising in the detection of new biomarkers. MALDI–TOF is an advanced mass spectrometry technique that allows for protein identification through its molecular characteristics, including protein sequence, structure, heterogeneity, cleavage, and post-translational modifications [25].

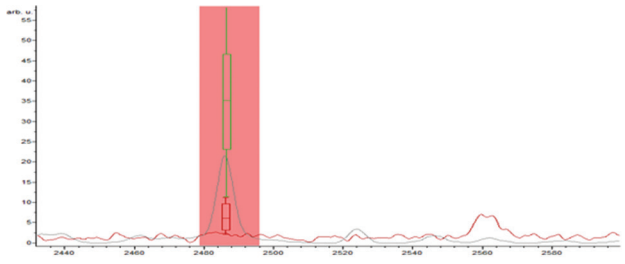
In the present study, WCX magnetic beads, UltrafleXtreme MALDI–TOF, and ClinProTools V 3.0 software were used to study the differential expression of peptides and proteins in the plasma of healthy-controls and PCa patients.

The choice of using serum sample versus plasma sample was a debatable issue in our proteomic analysis. The specimen committee of the Human Proteome Organisation (HUPO) and Plasma Proteome Project (PPP) prefers plasma to serum for the analysis of proteins under 20 kDa. The reason behind this is that plasma contains less degradation-produced peptides that are produced by the action of proteases during the coagulation process. However, the standardization of sample handling while processing is as important as the choice of type of the sample [26].

The results illustrated 26 peaks that showed significant difference between controls and patients. In total, 20 of those peaks were underexpressed, while six were overexpressed in the patients.

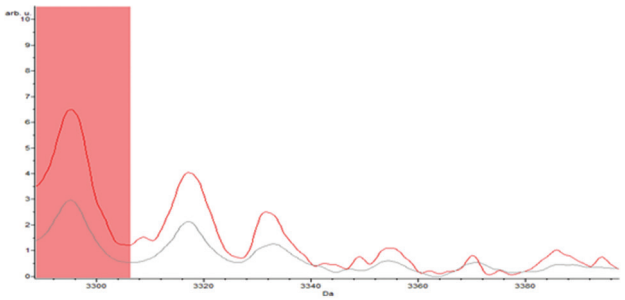
Petricoin *et al.* [27]. performed their study on benign and malignant PCa patients, and it showed the serum pattern that consisted of the combined relative amplitudes at seven *m/z* values that were 2092, 2367, 2582, 3080, 4819, 5439, and 18 220. With these results, 36 of 38 PCa

Figure 3



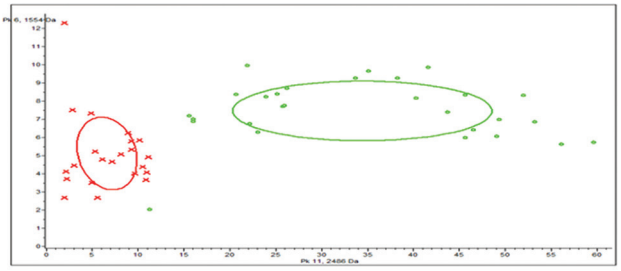
Peak 24 expression between patients and controls.

Figure 4



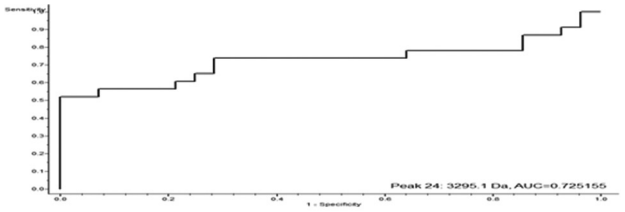
Box-and-whiskers command for peak 11 for the two classes.

Figure 5



2D peak statistics using the first two peaks. It shows the use of the first two peaks to separate the patients (red) from the controls (green).

Figure 6



Receiver-operating characteristic curve (ROC curve) for peak 24 with area under the curve (AUC)=0.725.

patients' proteomic pattern were correctly predicted (95%, 95% confidence interval = 82–99%), while 177 of 228 patients were accurately classified as holding benign conditions (78%, 95% confidence interval = 72–83%). For men with marginally elevated PSA levels (4–10 ng/ml, n = 137), the specificity was 71%.

It is the difference in the methodology between Petricoin *et al.*[27] and the current study that contributed to a difference in proteomic pattern. One difference was the usage of serum samples in their method rather than plasma. Moreover, they tested PCa against BPH and not against healthy participants. In addition, the work was performed using SELDI-TOF, which is often criticized for its reproducibility due to differences in the mass spectra obtained when using different batches of chip surfaces [28].

In another research study, Oh *et al.*[29] used serum samples from 179 PCa patients and 74 benign patients in their study. The samples were processed using ProXPRESSION Biomarker Enrichment Kits. Samples were subsequently analyzed by a MALDI/TOF mass spectrometer. For the selection of significant peaks out of the resulting peaks, a feature-selection algorithm called Extended Markov Blanket was used. In total, 26 peaks were achieved, including peaks with m/z 1274.3, 2244.3, 2425.3, 3432.3, 3947.1, and 4379.1, with an accuracy of 80.7%, a sensitivity of 83.5%, a specificity of 74.4%, a positive predictive value of 87.9%, and a negative predictive value of 68.2%. The samples were serum rather than plasma, which has been proven to be superior to serum in proteomics studies. The preparation technique used in this research study was more tedious than that of the current study, contributing to lesser peptide separation. Moreover, the 26 peaks detected did not coincide with the seven peaks from the study completed by Petricoin *et al.* [27].

Al-Ruwaili *et al.*[30] collected 99 serum samples from PCa patients in order to compare between aggressive and indolent types, and samples were tested using SELDI-TOF-mass spectrometry. The results identified 26 different peaks with a significant P value of less than 0.05. However, four peaks were then identified as candidate biomarkers with m/z of 9300 Da, which was upregulated in aggressive PCa patients and was identified as apolipoprotein C–I. Another three candidate biomarkers (22.2, 44.5, and 79.1 kDa) were found downregulated in the aggressive group, upregulated in the indolent group, and identified as apolipoprotein D, putative uncharacterized protein, and transferrin, respectively. Nevertheless, the discovered proteomic pattern differs from our research, as the comparison between different cancer stages was not undertaken in our study. The research merely conducted a comparison to detect new markers specific for the presence of cancer versus its lack thereof.

Although both the Al-Ruwaili *et al.*[30] and Petricoin *et al.*[27] studies were performed using serum samples and SELDI-TOF, their results did not coincide as they

were comparing two different patients (aggressive PCa vs. indolent/PCa vs. BPH). Moreover, the resulting 26 peaks from the Al-Ruwaili *et al.*[30] study differ from the 26 peaks in the study completed by Oh and colleagues, which shows that the difference in enrichment technique or mass spectrometer leads to different peak acquisition.

In the study by Ummanni *et al.* [31], tissue samples were tested from 24 cancer patients and their corresponding benign samples in 21 cases. Obtained tissues were processed using highly sensitive two-dimensional differential gel electrophoresis coupled with mass spectrometry. In total, 118 electrophoretic protein spots were revealed in cancer patients, which were later identified by mass spectrometer showing 79 proteins. In a similar investigation performed by Kuruma *et al.* [32], analysis of PCa proteomes using two-dimensional gel electrophoresis employing agarose gels for the initial isoelectric focusing step (agarose 2-DE) took place, with mass spectrometry used for protein identification. The discovered peaks had a higher mass range of more than 20 kDa. This was found to be exceeding the range tested in our research, as after 20 kDa, there were merely no peaks detected.

Some research studies used urine samples, as urine is considered to be an ultrafiltrate of plasma and its protein content was less viable to be disintegrated in comparison with other biological fluids [33]. In the study completed by Davalieva *et al.* [34], focus was on the identification of noninvasive biomarkers in urine with higher specificity than PSA. Urine samples from PCa and BPH patients were tested by two-dimensional differential gel electrophoresis coupled with mass spectrometry and bioinformatics analysis. In total, 23 proteins with differential abundance in urine of PCa patients compared with BPH patients were statistically significant ($P < 0.05$), out of which fibrinogen-A chain and inter-alpha-trypsin inhibitor heavy-chain 4 fragment (ITIH4) was upregulated, which coincided with peaks 12 (2660.7) and 11 (2485) in our study.

Theodorescu *et al.*[35] studied urine in 51 PCa and 35 patients with negative biopsy to identify a panel of polypeptides that could detect PCa. A panel consisting of 12 polypeptides was identified. One of them correlated with peak 6 (1553.97) in our study. The panel of biomarkers was validated in a blinded set of 213 samples (118 PCa and 95 negative biopsies). PCa was detected with 89% sensitivity and 51% specificity. This approach was tested for its effectiveness in routine clinical application in a subsequent study [36].

On comparing the achieved peak results with the database shown in the article done by Albrethsen [37], some peaks could be identified. Peak 3 (1061 Da) was identified as kininogen fragment, which was downregulated in patients in comparison with normal controls. The same peak was also mentioned in another article by Karpova *et al.* [38], but it was identified in bladder cancer patients. However, some explanations propose that there might be an overlap between the two cancers because of the common embryonic origin of both bladder and prostate. Additionally, peak 14 (2817 Da) – which was also downregulated – coincided with the mass of fibrinogen A1 fragment (2816 Da), while peak 11 of exact mass (2485) and mass range (2478–2496) was approximate to two peaks –ITIH4 of mass (2471 Da), and apolipoprotein A4 (2508 Da). The Apo-A4 chain was shown to be downregulated in the study by Fan *et al.*[39] Moreover, peak 19 with mass range (3228–3249) was identified as fibrinogen-A fragment, and was found to be upregulated in gray-zone patients versus other healthy controls.

Peak 24 with mass range 3288–3306 detected by the GA model was found to be approximate to ITIH4 fragment of mass (3272 Da), which was identified as a potential biomarker for the differentiation between PCa and BPH by Jayapalan *et al.* [40]. In addition, in our study, peak 23 with mass 3279.43 and mass range 3269.79–3288.51 was found to be significant in the univariate analysis with a PWKW of less than 0.016 and coincides with the mass of ITIH4 fragment. Peak 12 in the univariate analysis of mass 2660 Da was upregulated, it was found to be a significant peak according to its PWKW 0.00104 and was identified as fibrinogen-A fragment [41], a similar finding to what has been shown in the research by Davalieva *et al.* [34]. However, the marker was identified in urine.

In addition, other studies such as Schwamborn *et al.*[42] and Steurer *et al.*[43] performed their research using prostatic tissue rather than plasma samples. Nevertheless, Schwamborn *et al.*[42] processed their samples on the MALDI-imaging device and later used the ClinProTool program to perform their statistical analysis in concordance with the actual study. However, a 'support vector machine' algorithm was used to classify the cancerous from the noncancerous regions, achieving an overall cross-validation, sensitivity, and specificity of 88, 85.21, and 90.74% respectively. Four differing overexpressed peaks were discovered: 2753 and 6704 Da for noncancerous glands, and 4964 and 5002 Da for cancerous glands. The later peak was also found in the present study as peak 38 with mass 5002.59 Da; however, its PWKW of 0.095 made it illegible to be considered a significant peak.

As stated in the study performed by Malik *et al.* [44], the peak with *m/z* 8943 Da that was identified as

an isoform of apolipoprotein-AII was found to be upregulated in PCa patients compared with normal ones. This protein was found in several proteomic PCa studies. However, in our research, this increase was not detected, which should be searched for in future studies.

The field of proteomics has developed drastically during recent years. MALDI enables to identify and characterize the structure of proteins and peptides. Nonetheless, the main advantages of MALDI–TOF proteomic profiling are its high sensitivity, high thoroughput, ease of use, automation, and relatively low cost. It is beyond doubt that the development of further proteomic tools would aid in subsequent breakthroughs in cancer research.

The study might have been under the limitation of not being able to ascertain whether the discovered results were not because of the cancer pathology itself rather than PCa. The discovered peaks might have been attributed to prostatic pathologies rather than cancer such as BPH prostatitis.

Conclusion

Proteomic profiling can be an effective method for the discovery of new blood-based tumor markers for PCa patients. MALDI–TOF proteomic profiling represents a new frontier for screening and early diagnosis of PCa in Egypt. The analytical performance of multiplexed tumor profiles exceeds that of the single traditional tumor markers. However, the results that were found would need further validation on a larger scale and to be compared with other cancers to confirm its specificity for PCa rather than other inflammatory processes in the prostate or other cancers.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1 Smailova DS, Fabbro E, Ibrayev SE, Brusati L, Semenova YM, Samarova US, *et al*. Epidemiological and economic evaluation of a pilot prostate cancer screening program. Prostate Cancer 2020; 27:614062.

2 Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010; 127:2893.

3 Jemal A, Center MM, Ward E, Thun MJ. Cancer occurrence. Methods Mol Biol 2009; 471:3.

4 Kolonel LN, Altshuler D, Henderson BE. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. Nat Rev Cancer 2004; 4:519.

5 Ouyang B, Leung YK, Wang V, Chung E, Levin L, Bracken B, *et al*. alpha-Methylacyl-CoA racemase spliced variants and their expression in normal and malignant prostate tissues. Urology 2011; 77:249.

6 Wigle DT, Turner MC, Gomes J, Parent ME. Role of hormonal and other factors in human prostate cancer. J Toxicol Environ Health B Crit Rev 2008; 11:242.

7 Nelson WG, De Marzo AM, Isaacs WB. Prostate cancer. N Engl J Med 2003; 349:366.

8 Saad F, Clarke N, Colombel M. Natural history and treatment of bone complications in prostate cancer. Eur Urol 2006; 49:429.

9 Wolf AM, Wender RC, Etzioni RB, Thompson JM, D'Amico AV, Volk RJ, *et al*. American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA Cancer J Clin 2010; 60:70.

10 Catalona WJ, Smith DS, Wolfert RL, Wang TJ, Rittenhouse HG, Ratliff TL, *et al*. Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. JAMA 1995; 274:1214.

11 Adhyam M, Gupta AK. A review on the clinical utility of PSA in cancer prostate. Indian J Surg Oncol 2012; 3:120.

12 Coley CM, Barry MJ, Fleming C, Mulley AG. Early detection of prostate cancer. Part I: Prior probability and effectiveness of tests. The American College of Physicians. Ann Intern Med 1997; 126:394.

13 Loeb S, Roehl KA, Nadler RB, Yu X, Catalona WJ. Prostate specific antigen velocity in men with total prostate specific antigen less than 4 ng/ml. J Urol 2007; 178:2348.

14 Lee HY, Lee HJ, Byun SS, Lee SE, Hong SK, Kim SH. Classification of focal prostatic lesions on transrectal ultrasound (TRUS) and the accuracy of TRUS to diagnose prostate cancer. Korean J Radiol 2009; 10:244.

15 Duffy MJ. Role of tumor markers in patients with solid cancers: a critical review. Eur J Intern Med 2007; 18:175.

16 Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, *et al*. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology 1996; 14:61.

17 James P. Of genomes and proteomes. Biochem Biophys Res Commun 1997; 231:1.

18 Klopfleisch R, Klose P, Weise C, Bondzio., Multhaup G, Einspanier R, *et al*. Proteome of metastatic canine mammary carcinomas: similarities to and differences from human breast cancer. J Proteome Res 2010; 9:6380.

19 Grantzdorffer I, Carl-McGrath S, Ebert MP, Rocken C. Proteomics of pancreatic cancer. Pancreas 2008; 36:329.

20 Tiss A, Smith C, Camuzeaux S, Kabir M, Gayther S Menon U, *et al*. Serum peptide profiling using MALDI mass spectrometry: avoiding the pitfalls of coated magnetic beads using well-established ZipTip technology. Proteomics 2007; 7 Suppl: 1:77.

21 Eidhammer I, Flikka K, Martens L, Mikalsen SO. Computational methods for mass spectrometry proteomics. New Uork: John Wiley & Sons; 2008.

22 Bain BJ, Bates I, Laffan MA. Dacie and Lewis practical haematology. 12th ed.. China: Elsevier Health Sciences; 2016.

23 Wang H, Luo C, Zhu S, Fang H, Gao Q, Ge S, *et al*. Serum peptidome profiling for the diagnosis of colorectal cancer: discovery and validation in two independent cohorts. Oncotarget 2017; 8:59376.

24 Florkowski CM. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. Clin Biochem Rev 2008; 29(Suppl 1):S83.

25 Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem 2001; 70:437.

26 Tammen H, Schulte I, Hess R, Menzel C, Kellmann M, Mohring T, *et al*. Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. Proteomics 2005; 5:3414.

27 Petricoin EF, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, *et al*. Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst 2002; 94:1576.

28 Muthu M, Vimala A, Mendoza OH, Gopal J. Tracing the voyage of SELDI-TOF MS in cancer biomarker discovery and its current depreciation trend – need for resurrection? Trends Analyt Chem 2016; 76:95.

29 Oh JH, Lotan Y, Gurnani P, Rosenblatt KP, Gao J. Prostate cancer biomarker discovery using high performance mass spectral serum profiling. Comput Methods Programs Biomed 2009; 96:33.

30 Al-Ruwaili JA, Larkin SET, Zeidan BA, Taylor MG, Adra CN, Aukim-Hastie CL, *et al*. Discovery of serum protein biomarkers for prostate cancer progression by proteomic analysis. Cancer Genomics Proteomics 2010; 7:93.

31 Ummanni R, Mundt F, Pospisil H, Venz S, Scharf C, Barett C, *et al*. Identification of clinically relevant protein targets in prostate cancer with 2D-DIGE coupled mass spectrometry and systems biology network platform. PLoS ONE 2011; 6:e16833.

32 Kuruma H, Egawa S, Oh-Ishi M, Kodera Y, Maeda T. Proteome analysis of prostate cancer. Prostate Cancer Prostatic Dis 2005; 8:14.

33 Decramer S, Gonzalez de Peredo A, Breuil B, Mischak H, Monsarrat B, Bascands JL, *et al*. Urine in clinical proteomics. Mol Cell Proteomics 2008; 7:1850.

34 Davalieva K, Kiprijanovska S, Komina S, Petrosevska G, Zografska NC, Polenakovic M. Proteomics analysis of urine reveals acute phase response proteins as candidate diagnostic biomarkers for prostate cancer. Proteome Sci 2015; 13:2.

35 Theodorescu D, Schiffer E, Bauer HW, Duowes F Eichhorn F, Polley R, *et al*. Discovery and validation of urinary biomarkers for prostate cancer. Proteomics Clin Appl 2008; 2:556.

36 Schiffer E, Bick C, Grizelj B, Pietzker S, Schofer W. Urinary proteome analysis for prostate cancer diagnosis: cost-effective application in routine clinical practice in Germany. Int J Urol 2012; 19:118.

37 Albrethsen J. The first decade of MALDI protein profiling: a lesson in translational biomarker research. J Proteomics 2011; 74:765.

38 Karpova MA, Moshkovskii SA, Toropygin IY, Archakov AI. Cancer-specific MALDI-TOF profiles of blood serum and plasma: biological meaning and perspectives. J Proteomics 2010; 73:537.

39 Fan Y, Murphy TB, Byrne JC, Brennan L, Fitzpatrick JM, Watson RW. Applying random forests to identify biomarker panels in serum 2D-DIGE data for the detection and staging of prostate cancer. J Proteome Res 2011; 10:1361.

40 Jayapalan JJ, Ng KL, Razack AH, Hashem OH. Identification of potential complementary serum biomarkers to differentiate prostate cancer from benign prostatic hyperplasia using gel- and lectin-based proteomics analyses. Electrophoresis 2012; 33:1855.

41 Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, *et al*. Erdjument-Bromage H, Olshen AB, *et al*. Differential exoprotease activities confer tumor-specific serum peptidome patterns. J Clin Invest 2006; 116:271.

42 Schwamborn K, Krieg RC, Reska M, Jakse G, Knuechel R, Wellmann A. Identifying prostate carcinoma by MALDI-Imaging. Int J Mol Med 2007; 20:155.

43 Steurer S, Borkowski C, Odinga S, Buchholz M, Koop C, Huland H, *et al*. MALDI mass spectrometric imaging based identification of clinically relevant signals in prostate cancer using large-scale tissue microarrays. Int J Cancer 2013; 133:920.

44 Malik G, Ward MD, Gupta SK, Trosset MW, Grizzle WE, Adam BL, *et al*. Serum levels of an isoform of apolipoprotein A-II as a potential marker for prostate cancer. Clin Cancer Res 2005; 11:1073.