CANCER GENOMICS & PROTEOMICS 7: 93-104 (2010)

Discovery of Serum Protein Biomarkers for Prostate Cancer Progression by Proteomic Analysis

JAMAL A. AL-RUWAILI[1,2,3], SAMANTHA ET. LARKIN[1], BASHAR A. ZEIDAN[2], MATTHEW G. TAYLOR[2], CHAKER N. ADRA[3,4], CLAIRE L. AUKIM-HASTIE[1] and PAUL A. TOWNSEND[2]

[1]*School of Pharmacy and Biomedical Sciences, University of Portsmouth, Portsmouth, PO1 2DT, Hampshire, UK;*
[2]*Human Genetics Division, MP808, Southampton University General Hospital, University of Southampton, Southampton, SO16 6YD, Hampshire, UK;*
[3]*Stem Cell Therapy Programme, Research Centre, King Faisal Specialist Hospital and Research Centre, Riyadh, 11211, Kingdom of Saudi Arabia;*
[4]*Transplantation Center, Children's Hospital Boston and Brigham & Women's Hospital, Harvard Medical School, Boston, MA, U.S.A.*

Abstract. *Background: The incidence of prostate cancer (PCa) has increased in recent years due to the aging of the population and increased testing; however, mortality rates have remained largely unchanged. Studies have shown deficiencies in predicting patient outcome for both of the major PCa diagnostic tools, namely prostate specific antigen (PSA) and transrectal ultrasound-guided biopsy. Therefore, serum biomarkers are needed that accurately predict prognosis of PCa (indolent vs. aggressive) and can thus inform clinical management. Aim: This study uses surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS) mass spectrometry analysis to identify differential serum protein expression between PCa patients with indolent vs. aggressive disease categorised by Gleason grade and biochemical recurrence. Materials and Methods: A total of 99 serum samples were selected for analysis. According to Gleason score, indolent (45 samples) and aggressive (54) forms of PCa were compared using univariate analysis. The same samples were then separated into groups of different recurrence status (10 metastatic, 15 biochemical recurrences and 70 non-recurrences) and subjected to univariate analysis in the same way. The data from Gleason score and recurrence groups were then analysed using multivariate statistical analysis to improve PCa biomarker classification. Results: The comparison between serum protein spectra from indolent and aggressive samples resulted in the identification of twenty-six differentially expressed protein peaks ($p<0.05$), of which twenty proteins were found with 99% confidence. A total of 18 differentially expressed proteins ($p<0.05$) were found to distinguish between recurrence groups; three of these were robust with $p<0.01$. Sensitivity and specificity within the Gleason score group was 73.3% and 60% respectively and for the recurrence group 70% and 62.5%. Conclusion: SELDI-TOF-MS technology has facilitated the discovery of prognostic biomarkers in serum that can successfully discriminate aggressive from indolent PCa and also differentiate between recurrence groups.*

Correspondence to: Jamal. A Al-Ruwaili, School of Pharmacy and Biomedical Sciences, University of Portsmouth, Saint Michaels Building, Portsmouth, Hampshire, PO1 2DT, U.K. Tel: +44 2392845609, Fax: +44 2392843565, e-mail: Jamal.Alruwaili@port.ac.uk

Key Words: Biomarkers, mass spectrometry, prostate cancer, progression, serum proteomic analysis, SELDI-TOF-MS.

Prostate cancer (PCa) is the second most common cause of cancer in men in the Western world (1), resulting in 10,000 UK deaths per year (2). However, it is an unusual cancer in that a far more indolent form of the disease also exists in the majority of men aged over 80 years (3). At present, raised prostate-specific antigen (PSA) levels and a suspicious digital rectal examination (DRE) are used by clinicians to identify patients who require a transrectal ultrasound (TRUS)-guided biopsy of their prostate gland to look for PCa (4). However, PSA is neither particularly sensitive (5) nor specific (6). One shortcoming of PSA is that increases are also found in non-cancerous diseases such as benign prostatic hyperplasia (BPH), chronic inflammation or infection (7-9). PSA is also found in women (10). DRE is very subjective and even TRUS-guided biopsy is imperfect (11). This means that patients with aggressive forms of PCa may be missed and those with indolent disease may be subjected to unnecessary radical therapy and its associated morbidity. Therefore, new biomarkers that can diagnose PCa

with a higher sensitivity and specificity, and better predict prognosis to help guide management are greatly needed. Current research is therefore focused on using proteomic techniques to detect new biomarkers with higher specificity and sensitivity (12). Indeed, it is now moving towards using 'bio-signatures' made up of many biomarkers to test for disease, facilitated by novel proteomic technologies such as high-throughput mass spectrometry. One version of this technology is known as surface-enhanced laser desorption time-of-flight mass spectrometry (SELDI-TOF-MS) (13). SELDI-TOF-MS facilitates use of protein chip arrays with different binding affinities such as: metal affinity, IMAC-Cu; hydrophobic, WCX2, C16/H4 and weak cation exchange (CM10). The principle of this method is to extract proteins or peptides and apply them to surfaces enhanced with particular binding chemistries. These binding chemistries act to selectively bind proteins within the sample to form specific subsets. Unbound proteins are washed away and captured proteins are then ionized with SELDI and then measured by TOF-MS. Thousands of protein peaks are generated and displayed as spectra, map, or gel view, which can then be used to distinguish sample differences (14, 15). A computer algorithm then analyses and selects 'discriminatory' peaks (16). SELDI-TOF-MS has been proven to have a potential role in biomarker discovery and to be beneficial for early diagnosis, appropriate prognosis, effective therapy and monitoring of cancer treatment (17). Many studies have been published on PCa using SELDI-TOF-MS (18-25). However, only a limited number of studies have investigated PCa progression (26). This study uses SELDI-TOF-MS to identify differentially expressed serum proteins between PCa patients with indolent (Gleason score <7) and aggressive disease (Gleason score ≥7), and between patients with biochemical (BR) or metastatic disease (MR) recurrence and without disease recurrence (NR). Identification of progression-associated biomarkers would help clinicians individualise treatment at an early stage, which is particularly important in patients with a more aggressive form of the disease as there are no treatment options in the latter stages. Furthermore, increased knowledge regarding the course of disease may also limit unnecessary biopsy and surgical procedures which all have their associated risks. To our knowledge, this study represents the first to research candidate biomarkers from indolent and aggressive forms of prostate cancer using SELDI-TOF-MS technology.

Materials and Methods

Serum samples and ethical approval. PCa patients (n=99) were recruited at Queen Alexandra (QA) Hospital, Portsmouth between 2005 and 2008. Only samples from patients retaining their prostate at the time of sampling and with known Gleason score were used. Voluntary written informed consent for the research project was given by all participants and ethical approval for the collection of surplus serum was obtained (Approval number, 03/10/095). Disease was classified according to the histopathology-based Gleason score, with a Gleason score of <7 classed as indolent, and ≥7 classed as aggressive. This cut-off was selected as men with a Gleason score of 7 or above had a greatly increased risk of dying from the disease according to 20-year follow-up data (27). Ninety-nine serum samples from 45 indolent and 55 aggressive cases were used in the study. Samples were taken in a standardised manner using serum separator tubes (SST) for routine PSA testing. SST tubes contain a polymer gel and a clot activator (covering the inner wall of the tube). The sample was incubated for 30 minutes to allow the blood clot to form. It was then centrifuged at 1300-2000xg in 20-25˚C for 10 minutes. The supernatant (serum) was obtained by aspiration and placed into 5 ml labelled tube and stored at 4˚C. Following routine PSA tests, surplus material was stored for use in this trial. Upon collection, samples were aliquoted into 5×1.7 ml microcentrifuge tubes (50 μl from each patient sample) to minimise freeze/thaw during the study, and then stored at –20˚C.

Preparation of serum samples. Pre-analytical parameters were optimised using pooled samples to determine the optimum chip chemistry, buffer and sinapinic acid (SPA) concentration as well as SELDI analysis conditions. This showed that the CM10 chip (weak cation exchange) with 50 mM ammonium acetate buffer pH 4.5 (binding buffer) and 50% SPA in acetonitrile (ACN)/0.5% triflouroacetic acid (TFA) were optimal for this study (data not shown). Pooling an equal amount of serum from each tumour type (indolent and aggressive) served as a quality control (QC) for assessing the reproducibility and chip variability of the study. Serum samples were defrosted on ice for 90 minutes prior to use and then diluted with a denaturing buffer (U9, 9M Urea, 2% CHAPS, 50mM Tris-HCl pH 9) to generate a 1:10 dilution before being incubated on ice for 30 minutes. CM10 arrays were prepared using a BioMeK 3000 (Beckman Coulter) liquid handling robot. Arrays were pre-wetted with 5 μl of binding buffer for 5 minutes at room temperature, before two additions of 200 μl of binding buffer for 10 minutes at room temperature with shaking on a Micomix 5 platform shaker (DPC; form 20, amplitude 7). A total of 10 μl of U9-treated serum was then added to 90 μl of binding buffer and incubated for 60 minutes at room temperature with shaking. The arrays were then washed three times with 200 μl of binding buffer for 10 minutes each, followed by three HEPES (5 mM HEPES, pH 7) rinses to eliminate any unbound proteins or buffer salts which could interfere with array analysis. Prior to SELDI-TOF-MS analysis, the arrays were air dried for 45 minutes and then treated with two additions of 1 μl of 50% SPA in 50% TFA. All samples were applied to the chips in duplicate (to minimise drop out) and a QC sample was included on one spot of each chip from all arrays prepared.

SELDI-TOF-MS protein profiling. The Human Genetics Division proteomics laboratory at the University of Southampton provided the Enterprise 4011 SELDI-TOF-MS platform for PCa serum sample analysis. Protein Chip Data Manager v3.0.7 Software (PCDM) was used for analysis. A mass range between 0-10 kDa was examined with a focus mass of 5 kDa and matrix attenuation at 1 kDa, whereas a mass range above 10 kDa was studied with a 19 kDa focus mass and 5kDa matrix attenuation. External calibration of the SELDI-TOF-MS was carried out using a calibration kit of protein standards comprised of: recombinant hirudin (6.96 kDa), equine cytochrome *c* (12.23 k Da), equine

myoglobin (16.95 kDa), and carbonic anhydrase (29.0 kDa). Mass accuracy was found to be approximately 0.02% of actual mass value. Noise was eliminated using the noise definition setting which removes chemical noise resulting from the energy absorbing matrix in the low mass range. Therefore, only the area above matrix attenuation (1 kDa) was usable. Clusters were composed of peaks with a minimum signal-to-noise (S/N) ratio of 5 and valley depth of at least 5, using peaks which were present in at least 20% of the relevant spectra. The average mass of clusters was then used to identify unlabelled spectra, and therefore, each spectrum showed a peak intensity value. The mass window for each cluster was identified as 0.3% of the peak mass for low mass (0-10 kDa) and high mass range (>10 kDa) of the spectra optimised. The validity of peaks used as candidate biomarkers was confirmed by further evaluation and relabelling prior to final statistical analysis. Automatic detection was used to identify qualified mass peaks with a signal-to-noise ratio of more than five and a mass value range of 1-200 kDa. Further peak clusters were added after applying a second peak selection with (S/N of >2, within 0.3% mass window).

Data processing, normalisation, peak detection, feature selection and statistical analysis. Baselines were subtracted from the spectra which were then normalised within the mass range of 1-200 kDa. The normalisation factor was used to determine the outlier spectra showing possible variation caused by sample processing. Spectra with a normalisation factor of two-fold higher than the average were evaluated to assess quality. Any spectra with high background noise were deleted or replaced with duplicate spectra. Univariate analysis was then performed using Mann-Whitney *U*-tests to compare protein peak intensity between two groups. Ninety-nine serum samples from PCa patients with Gleason score and recurrence data at the time of this study were included. In terms of Gleason score, indolent (45 samples) and aggressive (54 samples) forms of PCa were compared using univariate analysis. The same samples were then separated into recurrence status groups (10 MR, 15 BR and 70 NR) and subjected to univariate analysis. For multivariate analysis, the mass (*m/z*) and intensity values for high and low mass ranges were combined and exported to build the tree algorithm. Multivariate statistical analysis was then used to classify the samples and evaluate the sensitivity and specificity using a tree classification algorithm. Multivariate analysis was performed using 69 (39 aggressive, 30 indolent) randomly assigned serum samples as a training set and 30 (15 indolent, 15 aggressive) as a test set. Multivariate analysis of recurrence was performed using 67 (10 BR, 7 MR, 50 NR) and 28 (5 BR, 3 MR, 20 NR) serum samples as training and test sets, respectively. Four samples were excluded from the recurrence analysis due to lack of recurrence information. To assess variation, the coefficient of variation (CV) of pooled QC samples were calculated using the formula:

$$CV=\sqrt{(CV_1^2 + CV_2^2 + CV_n^2)/n}$$

where: n=the number of peaks in the spectra.

Results

Serum candidate biomarkers categorised by Gleason score. Using the SELDI-TOF-MS software to study all spectra, 43 and 50 clusters of peaks were generated in the low mass and high mass ranges, respectively. The intensity of these clusters was compared between indolent and aggressive groups and yielded 26 statistically significant ($p<0.05$) differentially expressed peaks. These differentially expressed peaks included 18 underexpressed and 8 over-expressed biomarkers in the aggressive group relative to the indolent cohort (Table I). Examples of the differentially expressed protein peaks are shown in Figure 1. Interestingly, all of the underexpressed biomarkers occurred within the high mass range and all of the over-expressed biomarkers occurred within the low mass range. Twenty-two biomarkers were the most plausible ion peaks ($p\leq0.01$) with 99% confidence between the two groups. Exported peak data for the training and test sets were then evaluated by multivariate analysis and a classification tree with three nodes produced from the training group. Included in the tree were the peaks 4643 Da, 44583 Da and 4397 Da (Figure 2). The relative 'cost' of the model tree represents the heterogeneity of each node and was calculated to be 0.43 for this optimal tree. Two nodes were consistent

Table I. *SELDI protein peaks in serum discriminating the indolent group from the aggressive group. The intensity of clusters was compared between indolent and aggressive groups and yielded 26 statistically significant differentially expressed peaks. Twenty-two biomarkers were the most plausible ion peaks (p≤0.01) with 99% confidence between the two groups. The other four were significant at (p<0.05). Among protein peaks, 8 protein peaks were up regulated (↑) and 18 were down regulated (↓) in serum samples of patients with aggressive disease.*

Mean *m/z*	*p*-Value	Mean intensity		Expression change in aggressive disease
		Indolent	Aggressive	
4644.07	0.00056	1.53	2.20	↑
9298.54	0.00098	7.79	10.27	↑
2746.60	0.00295	1.66	2.99	↑
7770.51	0.00563	22.50	28.12	↑
3880.35	0.00613	2.77	3.47	↑
4615.73	0.00913	1.30	2.25	↑
3888.01	0.02197	2.44	2.99	↑
4396.25	0.03441	1.40	2.08	↑
44512.27	0.00019	2.82	2.27	↓
100608.45	0.00040	1.41	0.77	↓
28936.34	0.00052	4.10	2.45	↓
22226.88	0.00068	2.95	2.30	↓
14049.68	0.00073	16.82	10.97	↓
28101.22	0.00077	16.44	9.74	↓
89702.31	0.00087	1.28	0.68	↓
133986.69	0.00094	12.96	8.30	↓
111440.66	0.00139	0.36	0.27	↓
168304.09	0.00306	0.09	0.05	↓
95129.31	0.00486	0.70	0.42	↓
178854.19	0.00552	0.06	0.04	↓
141468.14	0.00552	1.90	1.54	↓
157657.89	0.00576	0.14	0.08	↓
118075.77	0.00756	0.10	0.07	↓
33315.15	0.00788	15.05	11.50	↓
79198.54	0.01155	3.71	2.91	↓
66685.96	0.01575	99.04	77.13	↓

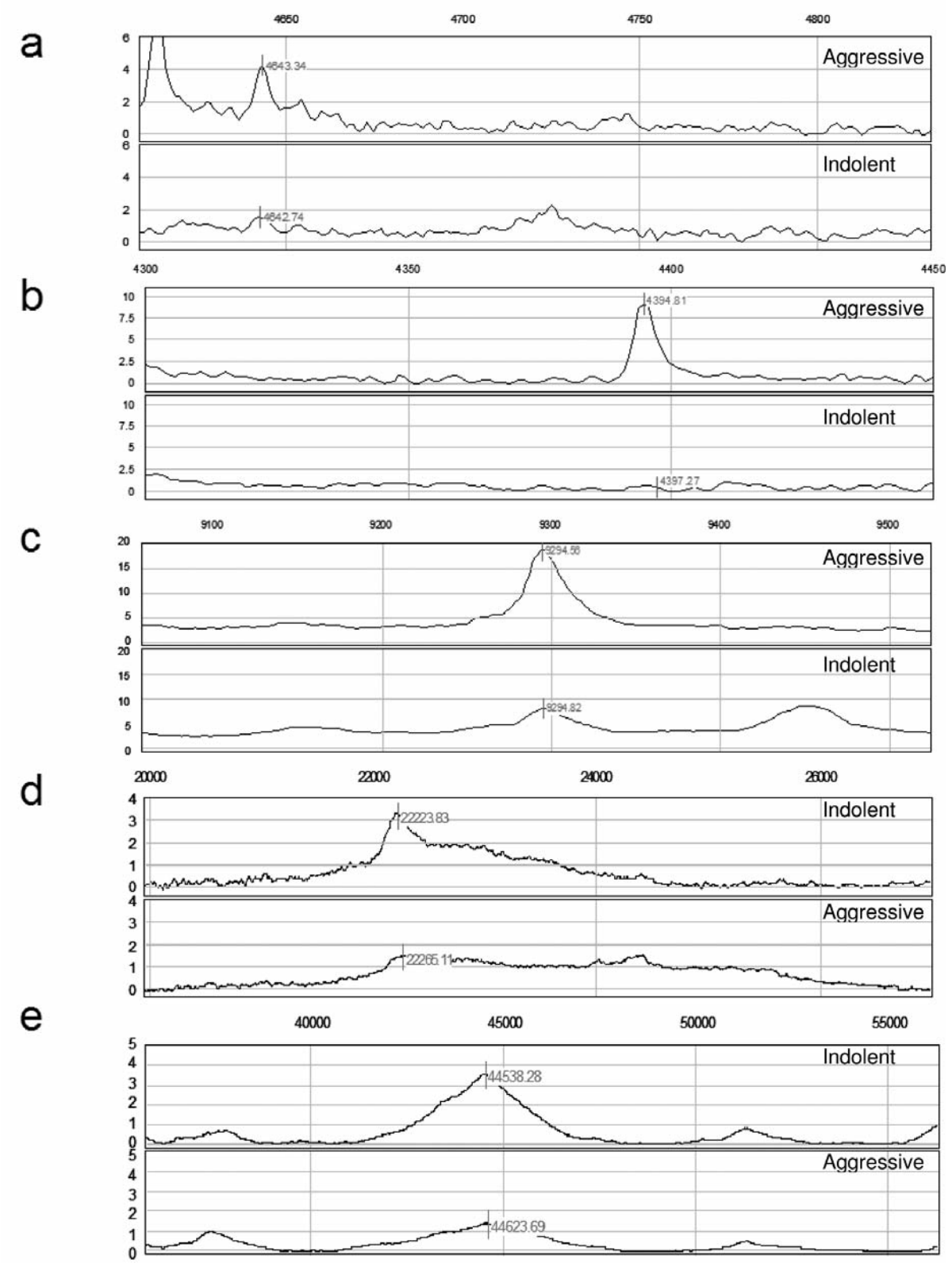

Figure 1. *Representative spectra of SELDI TOF-MS protein peaks: a: 4643 Da; b: 4397 Da; c: 9298 Da; d: 22226 Da and e: 44583 Da. Protein peaks a-c were overexpressed in aggressive disease compared to indolent disease, while those shown in d and e were overexpressed in the indolent group compared to the aggressive group. The X-axis shows the molecular weight; the Y-axis shows the relative peak intensity.*

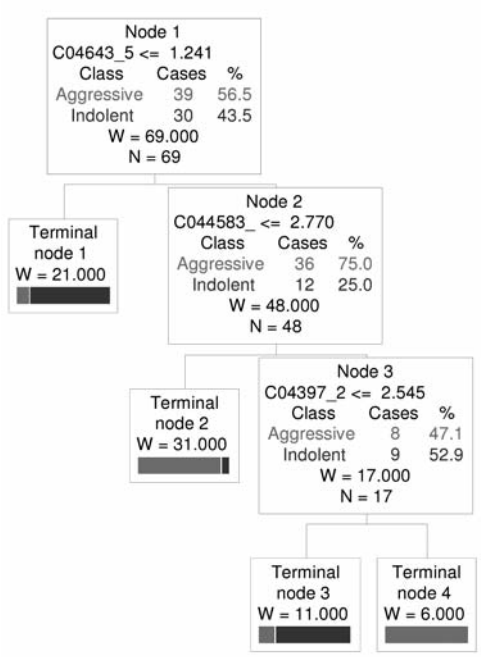

Figure 2. *Classification of patients with aggressive and with indolent PCa in the training set. A diagram of decision tree algorithm analyses explains the root node (top), the descendant nodes and the terminal nodes (Terminal node 1-Terminal node 4) as rectangles. Each node represents the classes (top number, number of aggressive samples; bottom number, number of indolent samples). The first number under the root and descendant nodes is the mass value followed by the peak intensity value. For example, the mass value under the root node is 4643 Da, and the intensity is ≤1.241. Three biomarkers were generated from the multivariate analysis. The biomarkers correspond to 4643, 44583 and 4397 Da.*

with differentially expressed proteins identified *via* the univariate analysis, although 44583 Da was not. However, a slightly smaller peak of 44512 Da was identified *via* the univariate analysis, hence this is likely to be the same protein. The optimal tree was then used to assign patients from the test set into groups. A receiver operator characteristics (ROC) curve was used to evaluate the tree produced from the training set. The area under the curve (AUC) for the ROC was calculated as 0.897. The tree was then applied to the independent test set and the sensitivity and specificity calculated as 73.3% and 60% respectively (Table II).

Serum candidate biomarkers categorised by recurrence status. Using the SELDI TOF-MS software to study all spectra, 43 and 50 peak clusters were generated in the low mass and high mass ranges respectively. Univariate analysis across the three groups (NR, BR and MR) was performed and yielded twelve putative biomarkers with p-values <0.05. Three of these were significant with 99% confidence ($p \leq 0.01$; Figure 3a, b). These biomarkers were able to differentiate between NR, BR and metastatic groups. Four biomarkers (6450 Da, 6650 Da, 3079 Da and 8941 Da) showed an increase in expression associated with BR, and by a decrease associated with metastatic disease in comparison to patients with NR disease (Figure 3a). Conversely, six proteins (7190 Da, 7482 Da, 8352 Da, 8135 Da, 4061 Da and 7929 Da) showed down-regulation with BR but a similar level of expression in metastatic disease compared to NR disease (Figure 3b). One protein peak (12869 Da) had a similar level of expression in NR and BR, with an increase associated only with metastatic disease (Figure 3c). Another protein (2747 Da) had a similar expression level in BR and metastatic disease with a lower expression level evident in NR disease (Figure 3c). Interestingly, the latter protein also showed an increased expression in aggressive disease based on Gleason score (Table III). Comparisons of NR and BR groups identified five putative biomarkers (6856 Da, 3888 Da, 3152 Da, 4466 Da and 16254 Da) that differed significantly (Figure 3d). These markers were not identified *via* the three group analysis above. Four of these (6856 Da, 3888 Da, 3152 Da and 4466Da) showed an increased expression in the BR group

Table II. *Validation of classification tree algorithm with 30 test samples. The classification tree algorithm was generated using CART software (Classification and Regression Trees). The nodes of the tree were then applied to the independent test set (15 aggressive vs. 15 indolent). The sensitivity and specificity were calculated as 73.3% and 60% respectively.*

Actual cases	Total cases	Percentage correct	Predicted	
			Aggressive N=17	Indolent N=13
Aggressive	15	73.33	11	4
Indolent	15	60.00	6	9

Table III. *Validation of classification tree algorithm with 28 test samples. The classification tree algorithm was generated using CART software (Classification and Regression Trees). The nodes of the tree were then applied to the independent test set (20 NR vs. recurrence groups BR and MR). The sensitivity and specificity were calculated as 70% and 62.5% respectively.*

Actual cases	Total cases	Percentage correct	Predicted	
			Aggressive N=17	Indolent N=13
NR	20	70.00	14	6
BR + MR	8	62.50	3	5

NR: No recurrence; BR biochemical recurrence; MR: metastatic recurrence.

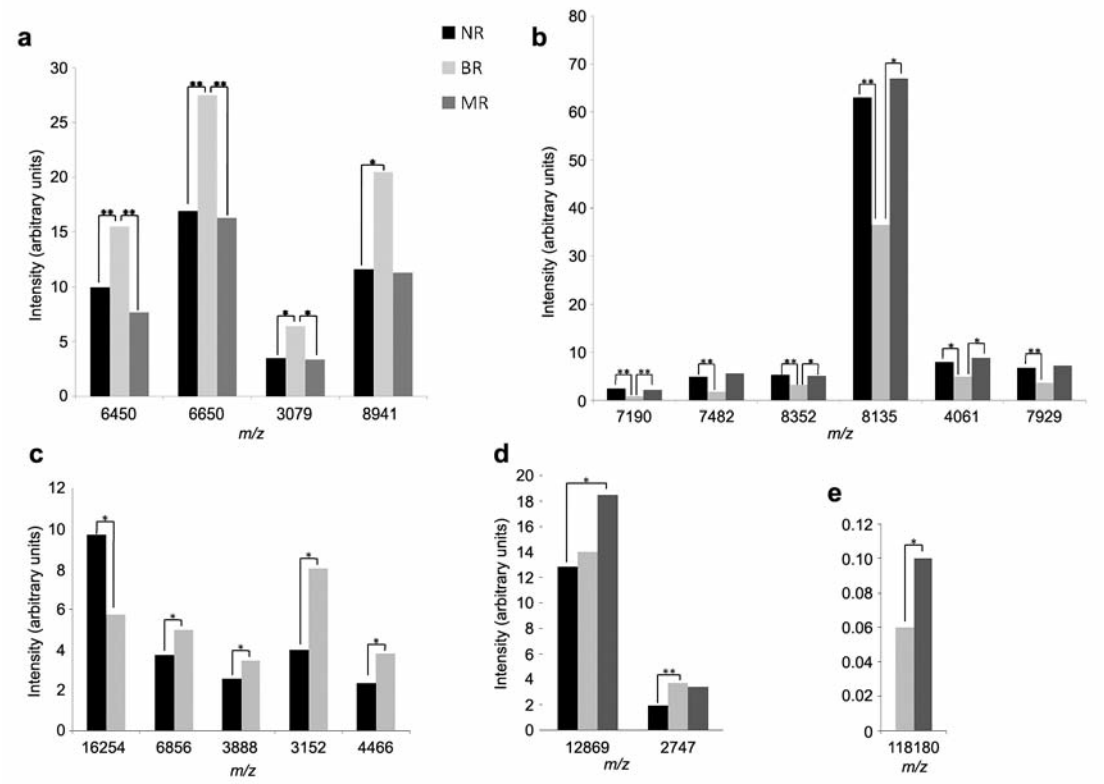

Figure 3. *SELDI protein peaks in serum discriminating the three recurrence groups (NR, BR and MR) represented by Gantt chart. **Indicates a difference at p≤0.01 with 99% confidence, while *indicates p<0.05 with 95% confidence. A total of 18 differentially expressed proteins were found to distinguish between different recurrence states. Three of these were significant with 99% confidence as shown in a and b. Four proteins from a (6450 Da, 6650Da, 3079 Da, 8941 Da) were found to be up regulated in BR, but with similar levels in NR and MR groups, and six proteins from b (7190 Da, 7482 Da, 8352 Da, 8135 Da, 4061 Da and 7929 Da) were lower in BR but with similar levels in NR and MR groups. c: One protein (12869 Da) was increased in metastatic disease but had similar levels in both NR and BR groups and a further protein (2747 Da) showed an increase in both BR and MR groups compared to NR. d: Comparison of NR and BR groups identified five markers. Four of these (6856 Da, 3888 Da, 3152 Da, 4466 Da) showed an increased expression in the BR group compared to the NR group, with only one (16254 Da) showing a decreased expression. e: Comparison between BR and MR groups produced one protein (118180 Da) which was found to be up-regulated in metastatic disease relative to BR.*

compared to the NR group, with only one (16254 Da) showing reduced expression. Furthermore, protein peak 3888 Da was also identified as being up-regulated in aggressive disease based on Gleason score. In addition to the markers identified from comparisons of all three recurrence groups, another marker was found to be significantly different between BR and MR groups. A protein peak of 118180 Da was found to show increased expression in metastatic disease compared to BR (Figure 3e). From the peak clusters, 67 serum samples from the training set and 28 from the test set were evaluated by multivariate analysis. Due to the limited numbers of samples from patients with recurrence, BR and MR groups were combined for multivariate analysis. A tree with 1 node (3314 Da) was produced from the training group (Figure 4) with a relative 'cost' of 0.45. The calculated AUC for the ROC was 0.771 and when validated using the test set, yielded a sensitivity and specificity of 70.0% and 62.5% respectively (Table III). However, this marker (3314 Da) was

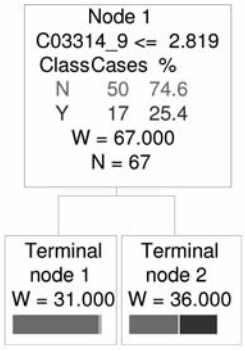

Figure 4. *Classification of non recurrence (N) and recurrence (Y) groups of PCa in the training set. A diagram of decision tree algorithm analyses shows the root node (top) and the terminal nodes (Terminal node 1-Terminal node 2). The node represents the classes (top number, number of non recurrence samples; bottom number, number of recurrence samples). A tree with one biomarker (3314 Da) resulted from the multivariate analysis.*

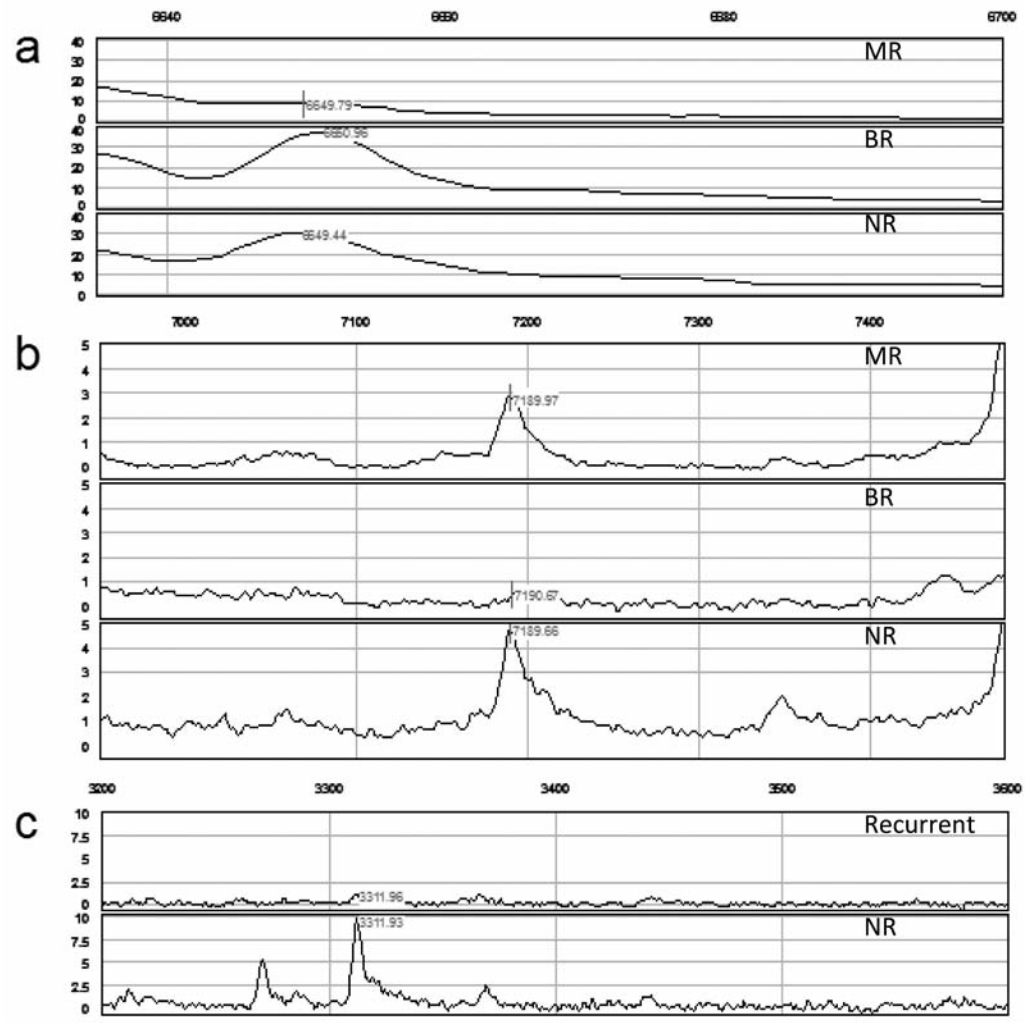

Figure 5. *Representative spectra of the peaks with a molecular mass of 6650 Da, 7190 Da, and 3314 Da obtained from SELDI-TOF-MS. a: A peak of 6650 Da with increased expression in BR and a decreased expression in MR and NR groups. b: A protein with mass 7190 Da was of low abundance in BR and high abundance in MR and NR groups. c: A protein of mass 3314 derived from multivariate analysis was found to be up-regulated in NR and down-regulated in recurrence groups. The X-axis shows the molecular weight; the Y-axis shows the relative peak intensity.*

not identified as being significantly differentially expressed by univariate analysis. Examples of the differentially expressed protein peaks are shown in Figure 5. An increased level of expression of 3314 Da was associated with recurrent disease (BR or MR; Figure 5c).

Discussion

The use of PSA as a diagnostic marker for PCa is not ideal due to its lack of specificity and sensitivity (5, 6). Many studies have tried to refine this by assessing PSA characteristics such as PSA density, PSA velocity, PSA doubling-time and age-specific PSA ranges (28-31), however these modifications have still failed to overcome the shortcomings of current PSA testing (5, 32). There is therefore a need to find new and better biomarkers for PCa that can lead to an earlier diagnosis and differentiate between indolent and aggressive forms of the disease (33).

Serum proteomic profiling has been vetted as a new diagnostic approach in PCa research. These profiles may be used to detect differences between malignant and benign disease and to differentiate between different subtypes of the same disease, such as indolent and aggressive forms of PCa (34). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) was the first technique developed for the identification of protein biomarkers from complex clinical samples and, until recently, gave the best snapshot of protein samples derived from cells (35, 36). However, 2D-PAGE is time-consuming, labour intensive and, more importantly, is unable to detect proteins smaller than 10,000 Da (37).

SELDI-TOF-MS overcomes this problem and has enabled the detection of this potentially important group of proteins (22, 38-41). The main advantages of the SELDI approach are its high throughput (42) ability to analyse proteins in body fluids directly on protein chip arrays without the necessity of long pre-analytical handling processes (43). Moreover, SELDI-TOF-MS technology can be applied to samples of low volume (1 μl) obtained from serum (44). It has been demonstrated that SELDI-TOF-MS has the potential role of obtaining high sensitivity, specificity and resolution for new biomarkers that are detected from different biological samples (45, 46). The analysis of samples using SELDI-TOF-MS may be easier than with other technologies as it does not require pathophysiological knowledge of the disease prior to analysis on the protein chip array (47). SELDI-TOF-MS has been utilised in many applications other than seeking new protein profiles, such as studies of protein–protein interactions (48, 49), protein–DNA interactions (50) and post-transitional modifications (51). However, SELDI faces hurdles that can limit its role in biomarker discovery. One of these obstacles is the reproducibility of protein chip arrays for samples that are analysed (52). The CV from array to array can influence the intensities of the peaks by10–40% (53). To overcome this variation, researchers contribute to the development of new methods to ensure the consistency of results. Research towards obtaining reproducible and standardised results is required if protein profiling derived from SELDI-TOF-MS is to become a diagnostic and prognostic tool (54). Furthermore, the high cost of SELD-TOF-MS may be prohibitive in widespread research (47).

In our study, the CM10 chip was adopted to analyse 99 serum samples from PCa patients. With masses between 2500 and 200000 Da, twenty-six differentially expressed proteins ($p<0.05$) were generated using Gleason score to classify patients as having aggressive or indolent disease: 18 biomarkers were underexpressed, and 8 were overexpressed in aggressive compared to indolent disease groups. From multivariate statistical data, a three-node tree was produced with an acceptable relative cost. The sensitivity and specificity of the model was 73.3% and 60% respectively when applied to the smaller test set. In terms of recurrence, 18 differentially expressed proteins ($p<0.05$) were found to distinguish between different recurrence states. Four proteins were found to be up-regulated in BR, but with similar levels in NR and MR groups, and six proteins were reduced in BR but with similar levels in NR and MR groups. One protein was increased in metastatic disease but had similar levels in both NR and BR, groups and a further protein showed an increase in both BR and MR groups compared to NR. The latter protein was also detected with Gleason score grouping analysis. Comparison of NR and BR groups identified five markers, four of which showed an increased expression in the BR group compared to the NR group, with only one showing decreased expression. In BR and MR group comparisons, one protein was found to show increased expression in metastatic disease relative to BR.

A one-node tree was produced from the training data which had 70% sensitivity and 62.5% specificity when tested. Deletion of this node results in trees with sensitivity and specificity reduced to 60% and 12.5%, respectively. This may indicate the importance of a using large sample size, particularly for the test set. A study by Pan *et al*. (55) used 178 serum samples to differentiate between PCa and a healthy control group. They found eight biomarkers from a tree classification algorithm with a sensitivity of 93% and specificity of 96% when applied to a test set. Six of these were consistent with the univariate analysis; the other two biomarkers were produced only in multivariate analysis, not with the univariate analysis. The deletion of the latter proteins resulted in lower sensitivity and specificity of the tree algorithm of 80% and 81%, respectively. This demonstrated the importance of the combined use of biomarkers for classification of disease cohorts. The limited number of samples from patients with recurrence may have diminished the ability of the tree classification algorithm to detect multiple markers for the tree and thus resulted in only one node which was not consistent with the univariate analysis. The CV for spectral reproducibility ranged from 27–38%. CVs of this and other studies are not directly comparable due to differences in sample type, protein chips and peak selection. However, CVs of this study are slightly lower than those of others (18, 56). Many studies (18, 22, 55, 57-59) have been performed to identify biomarkers that differentiate between PCa, BPH and healthy controls, but few have studied PCa progression. Indeed, there are only three such studies that have used SELDI-TOF-MS to study PCa progression, one of which utilised rat-derived cell lines (Dunning rat tumour cell R-3327) as a means of identifying molecular markers of progression (60), the second looked at differentiating between PCa patients with and without bone metastasis (19), and the third analysed markers of progression in the advanced hormonal stage of PCa relapse (61). These studies are not comparable with our own research because animal rather than human samples were used (60), or advanced stage disease was examined (19, 61). Our own study was categorised by Gleason score (<7 or ≥7) and recurrence status (NR, BR and MR).

The training and test sets require high numbers of samples to achieve reliable results. Adam *et al*. (18) used a training set composed of 167 PCa, 77 BPH, 82 healthy control cases and a test set with 30 PCA, 15 BPH and 15 healthy controls. The sensitivity and specificity were 83% and 97% respectively for PCA *vs*. healthy controls. Qu *et al*. (57) used a similar number of samples for their training and test sets (18). They identified diagnostic PCa models that were able to predict PCa with a sensitivity of a 97% and

a specificity of 97%. In our study, the sample size for training and test set was limited and therefore, a slightly lower sensitivity and specificity was achieved. Additionally, utilising multiple chip surfaces can result in a greater number of biomarkers and a more robust model. A study by Banez *et al*. (58) analysed samples using IMAC and CM10 chips. They analysed 44 PCa, 30 HC as a training set and 62 PCa and 26 healthy controls as test set. The sensitivity produced from CM10 chip was 63% and specificity was 77%. The sensitivity produced from IMAC chip was 34% and the specificity was 62%. However, 85% sensitivity and specificity was produced when IMAC and CM10 data were combined. The sensitivity and specificity for the model produced from the CM10 chip was similar to that produced by our study. This could be due to both their study and ours having similar training sample sizes.

In conclusion, this study represents the first investigation of candidate biomarkers differentiating indolent and aggressive forms of PCa based on SELDI-TOF-MS technology. The first, proof-of-concept stage of our research has alluded to proteomic differences which now require validation of the resultant biomarkers by immuno approaches.

Acknowledgements

We would like to thank Dr Sultan Al-Sadiri and our colleagues Professor Ian Cree, Mr Simon Holmes and Mrs Vanessa Basetter for their continued support and Dr Alan Cooper for his advice. We would like to thank Madeline fiji Schuck-Ranera and Mary Jane Anonuevo for their assistance. We would also like to thank the funding bodies for their commitment to our work: the Saudi Arabian Government (Ministry of Higher Education), and Wessex Medical Research.

References

1 Parkin DM, Bray F, Ferlay J and Pisani P: Global cancer statistics, 2002. CA Cancer J Clin *55*: 74-108, 2005.

2 http://www.cancerhelp.org.uk/help/default.asp?, Cancer Research UK. 2008. p. page=2656 used on 26/09/2008.

3 Bostwick DG, Burke HB, Djakiew D, Euling S, Ho SM, Landolph J, Morrison H, Sonawane B, Shifflett T, Waters DJ and Timms B: Human prostate cancer risk factors. Cancer *101*: 2371-2490, 2004.

4 Graham J, Baker M, Macbeth F and Titshall V: Diagnosis and treatment of prostate cancer: summary of NICE guidance. BMJ *336*: 610-612, 2008.

5 Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, Crowley JJ and Coltman CA Jr: Prevalence of prostate cancer among men with a prostate-specific antigen level < or = 4.0 ng per milliliter. N Engl J Med *350*: 2239-2246, 2004.

6 Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ, Crowley JJ, Parnes HL and Coltman CA Jr: Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. JAMA *294*: 66-70, 2005.

7 Roehrborn CG, McConnell JD, Lieber M, Kaplan S, Geller J, Malek GH, Castellanos R, Coffield S, Saltzman B, Resnick M, Cook TJ and Waldstreicher J: Serum prostate-specific antigen concentration is a powerful predictor of acute urinary retention and need for surgery in men with clinical benign prostatic hyperplasia. PLESS Study Group. Urology *53*: 473-480, 1999.

8 Roehrborn CG, McConnell J, Bonilla J, Rosenblatt S, Hudson PB, Malek GH, Schellhammer PF, Bruskewitz R, Matsumoto AM, Harrison LH, Fuselier HA, Walsh P, Roy J, Andriole G, Resnick M and Waldstreicher J: Serum prostate specific antigen is a strong predictor of future prostate growth in men with benign prostatic hyperplasia. PROSCAR long-term efficacy and safety study. J Urol *163*: 13-20, 2000.

9 Roehrborn CG, Boyle P, Gould AL and Waldstreicher J: Serum prostate-specific antigen as a predictor of prostate volume in men with benign prostatic hyperplasia. Urology *53*: 581-589, 1999.

10 Yu H and Berkel H: Prostate-specific antigen (PSA) in women. J La State Med Soc *151*: 209-213, 1999.

11 Selley S, Donovan J, Faulkner A, Coast J and Gillatt D: Diagnosis, management and screening of early localised prostate cancer. Health Technol Assess *1*: i, 1-96, 1997.

12 Pisitkun T, Johnstone R and Knepper MA: Discovery of urinary biomarkers. Mol Cell Proteomics *5*: 1760-1771, 2006.

13 Celis JE and Gromov P: Proteomics in translational cancer research: toward an integrated approach. Cancer Cell *3*: 9-15, 2003.

14 Paweletz CP, Trock B, Pennanen M, Tsangaris T, Magnant C, Liotta LA and Petricoin EF 3rd: Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. Dis Markers *17*: 301-307, 2001.

15 Sauter ER, Zhu W, Fan XJ, Wassell RP, Chervoneva I and Du Bois GC: Proteomic analysis of nipple aspirate fluid to detect biologic markers of breast cancer. Br J Cancer *86*: 1440-1443, 2002.

16 Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC and Liotta LA: Use of proteomic patterns in serum to identify ovarian cancer. Lancet *359*: 572-577, 2002.

17 Henderson NA and Steele RJ: SELDI-TOF proteomic analysis and cancer detection. Surgeon *3*: 383-390, 422, 2005.

18 Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer PF, Yasui Y, Feng Z and Wright GL, Jr: Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res *62*: 3609-3614, 2002.

19 Le L, Chi K, Tyldesley S, Flibotte S, Diamond DL, Kuzyk MA and Sadar MD: Identification of serum amyloid A as a biomarker to distinguish prostate cancer patients with bone lesions. Clin Chem *51*: 695-707, 2005.

20 Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, Cazares LH, Chan DW, Grizzle WE, Izbicka E, Kagan J, Malik G, McLerran D, Moul JW, Partin A, Prasanna P, Rosenzweig J, Sokoll LJ, Srivastava S, Thompson I, Welsh MJ, White N, Winget M, Yasui Y, Zhang Z and Zhu L: Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. Clin Chem *51*: 102-112, 2005.

21 Wagner M, Naik DN, Pothen A, Kasukurti S, Devineni RR, Adam BL, Semmes OJ and Wright GL Jr: Computational protein biomarker prediction: a case study for prostate cancer. BMC Bioinformatics *5*: 26, 2004.

22 Petricoin EF, 3rd, Ornstein DK, Paweletz CP, Ardekani A, Hackett PS, Hitt BA, Velassco A, Trucco C, Wiegand L, Wood K, Simone CB, Levine PJ, Linehan WM, Emmert-Buck MR, Steinberg SM, Kohn EC and Liotta LA: Serum proteomic patterns for detection of prostate cancer. J Natl Cancer Inst *94*: 1576-1578, 2002.

23 Martin DB, Gifford DR, Wright ME, Keller A, Yi E, Goodlett DR, Aebersold R and Nelson PSs: Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. Cancer Res *64*: 347-355, 2004.

24 Lehrer S, Roboz J, Ding H, Zhao S, Diamond EJ, Holland JF, Stone NN, Droller MJ and Stock RG: Putative protein markers in the sera of men with prostatic neoplasms. BJU Int *92*: 223-225, 2003.

25 Grizzle WE, Semmes OJ, Basler J, Izbicka E, Feng Z, Kagan J, Adam BL, Troyer D, Srivastava S, Thornquist M, Zhang Z and Thompson IM: The early detection research network surface-enhanced laser desorption and ionization prostate cancer detection study: A study in biomarker validation in genitourinary oncology. Urol Oncol *22*: 337-343, 2004.

26 You J, Cozzi P, Walsh B, Willcox M, Kearsley J, Russell P and Li Ys: Innovative biomarkers for prostate cancer early diagnosis and progression. Crit Rev Oncol Hematol, 2009.

27 Albertsen PC, Hanley JA and Fine Js: 20-year outcomes following conservative management of clinically localized prostate cancer. JAMA *293*: 2095-2101, 2005.

28 Carter HB, Pearson JD, Metter EJ, Brant LJ, Chan DW, Andres R, Fozard JL and Walsh PC: Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. JAMA *267*: 2215-2220, 1992.

29 Benson MC, Whang IS, Olsson CA, McMahon DJ and Cooner WH: The use of prostate specific antigen density to enhance the predictive value of intermediate levels of serum prostate-specific antigen. J Urol *147*: 817-821, 1992.

30 Oesterling JE, Jacobsen SJ and Cooner WH: The use of age-specific reference ranges for serum prostate specific antigen in men 60 years old or older. J Urol *153*: 1160-1163, 1995

31 Schmid HPs: Tumour markers in patients on deferred treatment: prostate specific antigen doubling times. Cancer Surv *23*: 157-167, 1995.

32 Thompson IM, Ankerst DP, Chi C, Goodman PJ, Tangen CM, Lucia MS, Feng Z, Parnes HL and Coltman CA Jr: Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. J Natl Cancer Inst *98*: 529-534, 2006.

33 Stamey TA, Caldwell M, McNeal JE, Nolley R, Hemenez M and Downs J: The prostate-specific antigen era in the United States is over for prostate cancer: what happened in the last 20 years? J Urol *172*: 1297-1301, 2004.

34 Alaiya AA, Roblick UJ, Franzen B, Bruch HP and Auer G: Protein expression profiling in human lung, breast, bladder, renal, colorectal and ovarian cancers. J Chromatogr B Analyt Technol Biomed Life Sci *787*: 207-222, 2003.

35 Pietrogrande MC, Marchetti N, Dondi F and Righetti PG: Decoding 2D-PAGE complex maps: relevance to proteomics. J Chromatogr B Analyt Technol Biomed Life Sci *833*: 51-62, 2006.

36 Weingarten P, Lutter P, Wattenberg A, Blueggel M, Bailey S, Klose J, Meyer HE and Huels C: Application of proteomics and protein analysis for biomarker and target finding for immunotherapy. Methods Mol Med *109*: 155-174, 2005.

37 Paweletz CP, Ornstein DK, Roth MJ, Bichsel VE, Gillespie JW, Calvert VS, Vocke CD, Hewitt SM, Duray PH, Herring J, Wang QH, Hu N, Linehan WM, Taylor PR, Liotta LA, Emmert-Buck MR and Petricoin EF 3rd: Loss of annexin 1 correlates with early onset of tumorigenesis in esophageal and prostate carcinoma. Cancer Res *60*: 6293-6297, 2000.

38 Seibert V, Wiesner A, Buschmann T and Meuer J: Surface-enhanced laser desorption ionization time-of-flight mass spectrometry (SELDI TOF-MS) and protein chip technology in proteomics research. Pathol Res Pract *200*: 83-94, 2004.

39 Wright GL Jr: SELDI proteinchip MS: a platform for biomarker discovery and cancer diagnosis. Expert Rev Mol Diagn *2*: 549-563, 2002.

40 Leak LV, Petricoin EF, 3rd, Jones M, Paweletz CP, Ardekani AM, Fusaro VA, Ross S and Liotta LA: Proteomic technologies to study diseases of the lymphatic vascular system. Ann NY Acad Sci *979*: 211-228; discussion 229-234, 2002.

41 Ardekani AM, Petricoin EF 3rd and Hackett JL: Molecular diagnostics: an FDA perspective. Expert Rev Mol Diagn *3*: 129-140, 2003.

42 Young DS and Tracy RP: Clinical applications of two-dimensional electrophoresis. J Chromatogr A *698*: 163-179, 1995.

43 Liotta LA, Ferrari M and Petricoin E: Clinical proteomics: written in blood. Nature *425*: 905, 2003.

44 Paweletz CP, Liotta LA and Petricoin EF 3rd: New technologies for biomarker analysis of prostate cancer progression: Laser capture microdissection and tissue proteomics. Urology *57*: 160-163, 2001.

45 Fung ET, Thulasiraman V, Weinberger SR and Dalmasso EA: Protein biochips for differential profiling. Curr Opin Biotechnol *12*: 65-69, 2001.

46 Yanagisawa K, Shyr Y, Xu BJ, Massion PP, Larsen PH, White BC, Roberts JR, Edgerton M, Gonzalez A, Nadaf S, Moore JH, Caprioli RM and Carbone DP: Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet *362*: 433-439, 2003.

47 Britten RA, Hardy C, Vlahou A, Gregory B, Giri PS and Drake R: Identification of reproducible low mass SELDI protein profiles specific to cisplatin resistance in human ovarian cancer cells. Oncol Rep *14*: 1323-1330, 2005.

48 Hinshelwood J, Spencer DI, Edwards YJ and Perkins SJ: Identification of the C3b binding site in a recombinant vWF-A domain of complement factor B by surface-enhanced laser desorption-ionisation affinity mass spectrometry and homology modelling: implications for the activity of factor B. J Mol Biol *294*: 587-599, 1999.

49 Stoica GE, Kuo A, Aigner A, Sunitha I, Souttou B, Malerczyk C, Caughey DJ, Wen D, Karavanov A, Riegel AT and Wellstein A: Identification of anaplastic lymphoma kinase as a receptor for the growth factor pleiotrophin. J Biol Chem *276*: 16772-16779, 2001.

50 Bane TK, LeBlanc JF, Lee TD and Riggs AD: DNA affinity capture and protein profiling by SELDI-TOF mass spectrometry: effect of DNA methylation. Nucleic Acids Res *30*: e69, 2002.

51 Thulasiraman V, Wang Z, Katrekar A, Lomas L and Yip TT: Simultaneous monitoring of multiple kinase activities by SELDI-TOF mass spectrometry. Methods Mol Biol *264*: 205-214, 2004.
52 Baggerly KA, Morris JS and Coombes KR: Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. Bioinformatics *20*: 777-785, 2004.
53 Lewczuk P, Esselmann H, Groemer TW, Bibl M, Maler JM, Steinacker P, Otto M, Kornhuber J and Wiltfang J: Amyloid beta peptides in cerebrospinal fluid as profiled with surface-enhanced laser desorption/ionization time-of-flight mass spectrometry: evidence of novel biomarkers in Alzheimer's disease. Biol Psychiatry *55*: 524-530, 2004.
54 Wong DT: Salivary diagnostics powered by nanotechnologies, proteomics and genomics. J Am Dent Assoc *137*: 313-321, 2006.
55 Pan YZ, Xiao XY, Zhao D, Zhang L, Ji GY, Li Y, Yang BX, He DC and Zhao XJ: Application of surface-enhanced laser desorption/ionization time-of-flight-based serum proteomic array technique for the early diagnosis of prostate cancer. Asian J Androl *8*: 45-51, 2006
56 Xiao Z, Adam BL, Cazares LH, Clements MA, Davis JW, Schellhammer PF, Dalmasso EA and Wright GL Jr: Quantitation of serum prostate-specific membrane antigen by a novel protein biochip immunoassay discriminates benign from malignant prostate disease. Cancer Res *61*: 6029-6033, 2001.
57 Qu Y, Adam BL, Yasui Y, Ward MD, Cazares LH, Schellhammer PF, Feng Z, Semmes OJ, and Wright GL, Jr: Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. Clin Chem 48: 1835-1843, 2002.
58 Banez LL, Prasanna P, Sun L, Ali A, Zou Z, Adam BL, McLeod DG, Moul JW and Srivastava S: Diagnostic potential of serum proteomic patterns in prostate cancer. J Urol *170*: 442-446, 2003
59 Li J, White N, Zhang Z, Rosenzweig J, Mangold LA, Partin AW and Chan DWs: Detection of prostate cancer using serum proteomics pattern in a histologically confirmed population. J Urol *171*: 1782-1787, 2004.
60 Malik G, Rojahn E, Ward MD, Gretzer MB, Partin AW, Semmes OJ and Veltri RW: SELDI protein profiling of dunning R-3327 derived cell lines: identification of molecular markers of prostate cancer progression. Prostate *67*: 1565-1575, 2007.
61 Kohli M, Siegel E, Bhattacharya S, Khan MA, Shah R and Suva LJ: Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) for determining prognosis in advanced stage hormone relapsing prostate cancer. Cancer Biomark *2*: 249-258, 2006.

Received November 18, 2009
Revised January 28, 2010
Accepted January 28, 2010