Discovery and validation of urinary biomarkers for prostate cancer

Dan Theodorescu1,2,3,* , Eric Schiffer4,* , Hartwig W. Bauer5, Friedrich Douwes6, Frank Eichhorn7, Reinhard Polley8, Thomas Schmidt9, Wolfgang Schöfer10, Petra Zürbig4, David M. Good11, Joshua J. Coon11,12, and Harald Mischak4

1Department of Urology, University of Virginia, Charlottesville, VA, USA
2Mellon Prostate Cancer Institute, University of Virginia, Charlottesville, VA, USA
3Department of Molecular Physiology, University of Virginia, Charlottesville, VA, USA
4Mosaaiques Diagnostics and Therapeutics AG, Hannover, Germany
5Ludwig-Maximilian-University Munich, Munich, Germany
6St. George Hospital Bad Aibling, Bad Aibling, Germany
7Urological Office, Bad Reichenhall, Germany
8Urological Office, Urologie am Aegi, Hannover, Germany
9Urological Office, Brackenheim, Germany
10Urological Office, Höchstadt, Germany
11Department of Chemistry, University of Wisconsin-Madison, Madison, WI, USA
12Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI, USA

Abstract

Only 30% of patients with elevated serum prostate specific antigen (PSA) levels who undergo prostate biopsy are diagnosed with prostate cancer (PCA). Novel methods are needed to reduce the number of unnecessary biopsies. We report on the identification and validation of a panel of 12 novel biomarkers for prostate cancer (PCA), using CE coupled MS. The biomarkers could be defined by comparing first void urine of 51 men with PCA and 35 with negative prostate biopsy. In contrast, midstream urine samples did not allow the identification of discriminatory molecules, suggesting that prostatic fluids may be the source of the defined biomarkers. Consequently, first void urine samples were tested for sufficient amounts of prostatic fluid, using a prostatic fluid indicative panel (“informative” polypeptide panel; IPP). A combination of IPP and PCA to predict positive prostate biopsy was evaluated in a blinded prospective study. Two hundred thirteen of 264 samples matched the IPP criterion. PCA was detected with 89% sensitivity, 51% specificity. Including age and percent free PSA to the proteomic signatures resulted in 91% sensitivity, 69% specificity.

© 2008 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
Correspondence to: Harald Mischak.
Correspondence: Dr. Harald Mischak, Mosaaiques Diagnostics and Therapeutics AG, Mellendorfer Str. 7-9, 30625 Hannover, Germany
E-mail: mischak@mosaiques-diagnostics.com Fax: +49-511-55474431.
*Both authors contributed equally.
Publisher’s Disclaimer: The authors have declared a conflict of interest. H. Mischak is founder and co-owner of Mosaaiques Diagnostics, which developed the CE-MS technology and the MosaaiquesVisu software. E. Schiffer and P. Zürbig are employees of Mosaaiques Diagnostics.
1 Introduction

The worldwide incidence of prostate cancer (PCa) ranks third among cancers in men [1]. The use of prostate specific antigen (PSA) has led to a significant increase in the number of diagnosed cases [1]. However, PSA is associated with significant false positives [2] and this results in up to 700 000 unnecessary prostate biopsies per year in the United States [3]. In addition, the Prostate Cancer Prevention Trial (PCPT), a phase III randomized, double-blind, placebo-controlled trial of finasteride for the prevention of PCa [4], indicated that PCa could be found at all levels of PSA and that even 15% of men with PSA<4 ng/mL had PCa [5]. These data highlight the urgent need for additional noninvasive biomarkers for PCa that may enable improvement both in the sensitivity and especially in the specificity of the current noninvasive methods used. The limitations have prompted serum autoantibody [6] or proteomic approaches as novel solutions to this diagnostic challenge [7-9]. However, the majority of studies analyzed either blood in form of serum or plasma, a body fluid which undergoes massive degradation upon sampling, thereby greatly impeding comparative analysis [10], or applied SELDI, a platform for which comparability of datasets is not easy to achieve [11-17]. We have developed CE coupled MS (CE-MS) [18] for the analysis of body fluids, with emphasis on urine. Urine was chosen as the body fluid of highest interest as it generally does not undergo proteolytic degradation after sampling, hence was found to be very stable for proteome analysis [19,20]. Further, urine has been shown to contain an array of disease-specific information, which can be assessed using proteome analysis, as also outlined in several recent reviews [21,22]. CE-MS analysis enabled the identification of biomarkers for several chronic renal diseases [23, 24], transplantation-associated complications [25,26], but also urogenital malignancies [19]. As CE-MS allows for the rapid (~60 min), robust, and reproducible analysis of several hundred urinary polypeptides from a small volume of urine, it can in fact be used not only as a discovery and validation platform but also as an application platform for diagnostic purposes. The CE-MS technology has been validated in routine clinical settings [25].

Based on our previous experiences, we hypothesized that urine may also contain valid biomarkers for PCa. We therefore initiated a study to identify urinary biomarkers for PCa and their subsequent validation in a blinded prospective study. The study was performed in accordance with the recently proposed guidelines for clinical proteomics [27], all data and metadata were collected as recently suggested by the MIAPE guidelines [28].

2 Materials and methods

2.1 Patients, procedures, and tumor demographics

2.1.1 “Training set” for biomarker discovery (N = 86)—From May 2004 to March 2005, patients scheduled to have a transrectal ultrasound (TRUS) guided biopsy due to elevated PSA and/or suspicious digital rectal examination (DRE) were offered participation in the current study. Patients provided the initial 10 mL of voided urine immediately prior to DRE and an 8-12 core prostate biopsy. Pathology was recorded as either PCa, high grade prostatic intraepithelial neoplasia (HGPIN), atypical small acinar proliferation (ASAP), or benign tissue, including chronic, and acute prostatitis [29,30]. Only patients with PCa (N = 51) and benign tissue (N = 35) were included in the study. Patient characteristics for the 86 patients of the training set are given in Table 1.
2.1.2 “Control set” for informative polypeptide panel (IPP) derivation \( (N = 184) \)

Midstream urine samples were prospectively collected at the University of Virginia (UVA) and Hannover Medical School (MHH) from healthy volunteers and patients with genitourinary diseases without known or suspected (abnormal DRE or PSA>1) PCa [19]. These were used to establish an IPP aimed at identifying a polypeptide panel characteristic of initial urine samples collected in the 86 patients described above. This approach was taken because we hypothesized that the first 10 mL of the urinary void would wash out most of the seminal/prostatic fluid present in the prostatic urethra, which would be the material most likely to contain the polypeptide panel differentiating PCa from benign disease [31,32]. These 184 samples included male individuals \( (N = 138, \text{mean age } 49 \pm 12 \text{ years } \pm \text{SD}) \) with no history or suspicion of PCa as well as females \( (N = 46, \text{mean age } 42 \pm 15 \text{ years } \pm \text{SD}) \). Female samples were used as negative controls for polypeptides that are expected to originate from seminal/prostatic fluid.

2.1.3 “Test set” for blinded evaluation of polypeptide panels diagnostic for PCa \( (N = 264) \)

The polypeptide panel developed from patients in the training set, which also had the informative panel, was evaluated in a blinded prospective study to determine its predictive ability for a positive prostate biopsy. From April 2005 to December 2006, 264 patients were accrued at 22 institutions, including two universities \( (N = 221) \) and 20 private medical offices \( (N = 43) \). All patients had elevated PSA \( (2<\text{PSA}<20) \) and/or suspicious DRE and underwent 8-12 core TRUS guided biopsy. Prior to DRE and biopsy, they provided the first 10 mL of voided urine. Patient characteristics for the 264 patients of the test set are given in Table 1.

2.1.4 Protection of human subjects—This study was approved by the Human Investigation Committees at the University of Virginia and Hannover Medical School. All participants gave written informed consent.

2.2 Urine collection

Voided urine was obtained from every patient who agreed to participate in the study, without any preselection. Urine samples for the 86 patients in the training set and the 264 in the validation set were taken before DRE and/or before prostate biopsy. Midstream urine samples from patients without PCa were collected without any prior knowledge as to whether the patient had a DRE before the collection. Samples were aliquoted and stored at -20°C, shipped on cold packs, and thawed only once before processing. The impact of storage and shipping on sample stability and other parameters has been described in detail [19].

2.3 Sample preparation

For proteomic analysis, a 0.7 mL aliquot of urine was thawed immediately before use and diluted with 0.7 mL of 2 M urea and 10 mM \( \text{NH}_4\text{OH} \) containing 0.02% SDS. In order to remove proteins of higher molecular mass, such as albumin or IgG, the sample was filtered using Centrisart ultra-centrifugation filter devices \( (20 \text{kDa MWCO}; \text{Sartorius, Goettingen, Germany}) \) at 3000 rcf until 1.1 mL of filtrate was obtained. The filtrate was then applied onto a PD-10 desalting column \( (\text{Amersham Bioscience, Uppsala, Sweden}) \) equilibrated in 0.01% \( \text{NH}_4\text{OH} \) in HPLC-grade \( \text{H}_2\text{O} \) \( (\text{Roth, Germany}) \) to decrease matrix effects by removing urea, electrolytes, and salts, and to enrich polypeptides present in the sample. Finally, all samples were lyophilized, stored at 4°C, and suspended in HPLC-grade \( \text{H}_2\text{O} \) shortly before CE-MS analysis [19,33].

2.4 CE-MS analysis

For proteomic analysis samples were prepared as previously described [19,33]. Briefly, CE-MS analysis was performed using a P/ACE MDQ CE system \( (\text{Beckman Coulter, Fullerton}) \).
CA, USA) on-line coupled to a Micro-TOF MS (Bruker Daltonic, Bremen, Germany). The performance of the sample preparation procedure as well as the analytical performance of the instrumental setup was evaluated. The average recovery of the sample preparation procedure is approximately 85% with a detection limit of ~1 fmol. The monoisotopic mass signals could be resolved for \( z \leq 6 \). The mass accuracy of the CE-TOF-MS method was determined to be <25 ppm for monoisotopic resolution and <100 ppm for unresolved peaks (\( z > 6 \)). The precision of the analytical method was determined by assessing (i) the reproducibility achieved for repeated measurement of the same aliquot and (ii) by the reproducibility achieved for repeated preparation and measurement of the same urine sample. The 200 most abundant polypeptides were detected with a rate of 98%. The performance of the analytical system over time was assessed with consecutive measurements of the same aliquot over a period of 24 h. No significant loss of polypeptides was observed implying the stability of the CE-MS set up, the post preparative stability of the urine samples at 4°C and their resistance to, e.g., oxidizing processes or precipitation.

2.5 Data processing and cluster analysis

Mass spectral ion peaks representing identical molecules at different charge states were deconvoluted into single masses using MosaiquesVisu software [19]. In addition, the migration time and ion signal intensity (amplitude) were normalized using internal polypeptide standards [34]. The resulting peak list characterizes each polypeptide by its molecular mass (kDa), normalized migration time (min), and normalized signal intensity. All detected polypeptides were deposited, matched, and annotated in a Microsoft SQL database, allowing further analysis and comparison of multiple samples (patient groups). Polypeptides within different samples were considered identical if the mass deviation was less than 100 ppm [19] and the migration time deviation was less than 3%. CE-MS data of all samples can be provided by request.

2.6 Statistical methods

Estimates of sensitivity and specificity were calculated based on tabulating the number of correctly classified samples [33]. Confidence intervals (95% CI) were carried out in MedCalc version 8.1.1.0 (MedCalc Software, Mariakerke, Belgium, http://www.medcalc.be). The receiver operating characteristic curve (ROC) was obtained by plotting all sensitivity values (true positive fraction) on the \( y \)-axis against their equivalent (1-specificity) values (false positive fraction) for all available thresholds on the \( x \)-axis (MedCalc Software). The area under the ROC curve (AUC) was evaluated, as it provides a single measure of overall accuracy that is not dependent upon a particular threshold [35]. Biomarker candidates were statistically analyzed using nonparametric methods, such as Wilcoxon’s test (rank sum test) with \( p < 0.05 \) as significance level. PCa specific biomarkers were additionally validated using the Benjamini-Hochberg procedure to control false discovery rate in multiple testing [36].

2.7 Definition of the IPP

To enable informative biomarker selection, only polypeptides with a frequency >67% in at least one group (training set [37] \( N = 86 \) or control set (male midstream) set \( N = 138 \)) were retained for further analysis. The discriminatory potential of polypeptides to distinguish initial and midstream urine samples was analyzed by ROC. The amplitude distribution of the CE-MS data of polypeptides present in the samples was used as the ROC variable and the affiliation to a diagnostic group (training vs. control) as the classification variable. The obtained AUC value of the analysis of a given polypeptide was interpreted as a measure of its discriminatory potential. Polypeptides were further validated by excluding those that were present with high frequency in female urine samples and congruously can be excluded to originate from seminal/prostatic fluid. All remaining polypeptides with AUC\( \geq 0.750 \) were statistically analyzed using...
nonparametric methods, such as Wilcoxon's test (rank sum test) with $p<0.05$ as significance level.

2.8 Definition of the PCa specific polypeptide panel (PCaP)

In order to enable biomarker definition, only polypeptides with a frequency >70% (459 of 1459 polypeptides) in at least one diagnostic group (training set $N=86$ (PCa $N=51$ or NED $N=35$)) were retained for further analysis. The discriminatory potential of polypeptides to distinguish PCa samples from those without any evidence of disease (NED) was analyzed using ROC curves. All polypeptides with AUC$\geq0.600$ (49 of 1459 polypeptides) were statistically analyzed using Benjamini and Hochberg procedures for multiple testing corrections [36]. Initial statistical analyses indicated that the PCa samples of the training set were heterogeneous. To address this heterogeneity, an additional bootstrapping process for the definition of additional PCa specific biomarkers was performed: The PCa samples of the training set were classified with initially defined biomarkers (Table 2B, not highlighted with asterisk) Those PCa samples that were not correctly classified in total crossvalidation procedures were compared once again with the NED samples. The markers defined in this additional step (Table 2B, highlighted with asterisk) were combined with the initial markers to create the final marker list (Table 2B).

2.9 Support vector machine (SVM) modeling and sample classification

The IPP and PCaP classification models were established using the training set and MosaCluster software package [33]. This software tool generates a model for each of the different groups investigated based on the polypeptides which are best suited to discriminate case and control groups. Each polypeptide used for classification represents one dimension in an $n$-dimensional space. The software allows the classification of samples in the high dimensional parameter space by utilizing SVMs. The result of SVM classification is a dimensionless number representing the Euclidian distance of data points to the separating hyper plane ($F$-factor). These algorithms already showed good performance in the evaluation of multidimensional data [38-40]. In order to classify the blinded test set, in a first step the initial urine samples were compared with the IPP model to assess for characteristic initial urine signatures. Those samples that showed sufficient match and were classified as informative were subsequently analyzed for PCa specific biomarker signatures (Fig. 1).

2.10 Nomogram development

Only those samples of the test set (Table 1A) for which information on percent free serum PSA (F/T-PSA) was available (in addition to serum PSA (T-PSA), patient age, and TRUS prostate volume (PV)) were used for nomogram development. The cohort included 109/264 urine samples (60/118 PCa and 49/95 NED). Statistical tests were performed using MedCalc version 8.1.1.0. Two-sided tests with significance at 0.05 were used. For the nomogram, the association between predictors and PCa on needle biopsy was tested in multivariate logistic regression. Therefore, significant predictors were entered sequentially. After entering a variable, those which then became nonsignificant were removed. The probability of PCa on prostate biopsy was calculated according to $p=1/1+e^{-\left(\beta_0+\sum x_i\beta_i\right)}$, with $\beta_0$ as constant and $\beta_i$ as the classification coefficient of the predictor $x_i$. Predictive accuracy of the obtained nomogram was quantified with ROC statistics (Fig. 2).

2.11 Sequencing of polypeptides

Candidate biomarkers and other native peptides from urine were sequenced using CE- and LC-MS/MS analysis as described [41]. Additional MS/MS experiments were performed on an Ultimate 3000 nanoflow system ( Dionex/LC Packings, USA) connected to an LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a
nanoelectrospray ion source. Chromatographic separation of the peptides was accomplished on a 10 cm fused-silica microcolumn of 75-μm inner diameter packed with RP Biosphere C18, 5 μm resin (NanoSeparations, Netherlands). The sample was injected onto the column at a flow rate of 5 μL/min and subsequently split to an elution rate of 250 nL/min using a linear gradient (60 min) from 2 to 50% CH₃CN in water (0.1% formic acid). The mass spectrometer was operated in data-dependent mode to automatically switch between MS and MS/MS acquisition. Full scan MS spectra (from m/z 300 to 2000) were acquired in the Orbitrap with resolution $R = 60,000$ at m/z 400 (target value of 500 000 charges in the linear IT). The five most intense precursor-ions were sequentially isolated for fragmentation in the linear IT using CID, and the detection took place either in the linear IT (parallel mode; target value 10 000) or in the Orbitrap (target value of 500 000). Orbitrap MS/MS spectra were acquired with resolution $R = 15,000$ at m/z 400. General mass spectrometric conditions were: electrospray voltage, 1.6 kV; no sheath and auxiliary gas flow; ion transfer tube temperature, 225°C; collision gas pressure, 1.3 mTorr; normalized collision energy, 32% for MS/MS. Ion selection threshold was 500 counts for MS/MS. Samples were also analyzed using Electron Transfer Dissociation (ETD) [42,43]. Briefly, samples were resuspended (50 μL of 100 mM acetic acid), loaded onto a 360× 75 μm microcapillary precolumn, and connected to a 360× 50 μm analytical column with a ~1 μm tip pulled with a laser puller (both columns were packed in-house with ~5-8 cm of C8 resin). Peptides were separated by nRP-HPLC (Agilent 1100; flow split by tee to ~60 nL/min) and introduced into an ETD-capable Finnigan LTQ quadrupole linear IT via nESI. The resulting MS/MS data were submitted to MASCOT (www.matrixscience.com) for a search against human entries in the MDSB Protein Database. Accepted parent ion mass deviation was 50 ppm and accepted fragment ion mass deviation was 500 ppm. Alternatively, data files were searched against the NCBI human nonredundant database using the Open Mass Spectrometry Search Algorithm (OMSSA), using an e-value cut-off of 0.01. All matched sequences were manually validated as described by Zürbig et al. [41].

3 Results

Data from our pilot study suggested that urinary polypeptides might contain biomarkers suitable for distinguishing between patients with and without PCa on biopsy [34]. In this study, we initially wanted to validate these findings in a blinded prospective study. However, the defined biomarkers failed to distinguish between malign and benign samples with sufficient accuracy (data not shown). In general, the test displayed increasing numbers of false negative rates. During the quest for explaining these observations we found several publications suggesting polypeptides originating from prostatic fluid as potential biomarkers for PCa [44-46]. We concluded that absent prostatic fluid in several of the collected midstream urine samples might have resulted in false negative results, since the PCa specific biomarkers originating form the prostatic secretions were missing or extremely diluted. Since prostatic fluid is preferentially present in the first voided urine [31,32], we decided to focus on this urine fraction as the assay fluid for a relaunched study. In this work, we defined PCa biomarkers specifically present in first void urine and evaluated the diagnostic performance of these potential biomarkers in a blinded prospective multicenter study of patients undergoing prostate biopsy. Our study also utilized extensive controls from patients with nonmalignant genitourinary diseases to establish quality control criteria for the samples and used different clinical centers to generalize the results as proposed standards [27].

The CE-MS data on all relevant polypeptides (present in >40% of the samples of one group) of all individual samples used in this study are given in Supporting Information Table 1. We compared the polypeptide profile of the collected first void urine to that of midstream urine samples to define initial void specific (prostatic secretion specific) polypeptide signatures. A comparison of the 86 first void urine samples of the training set (Table 1) to a set of 138 male midstream urine control samples (Table 1) was performed to define polypeptides that were
present in the first void urine samples, but generally not (or with substantially lower abundance) in the midstream urine samples (Table 1). Only markers that could be detected in at least two-thirds of the samples in a diagnostic group (e.g., detection frequency in initial urine samples) were analyzed using ROC statistics. Furthermore, a set of 46 female urine samples was used to exclude any nonseminal/prostatic fluid specific polypeptides. Those polypeptides with AUC values >0.75 were subsequently statistically analyzed. A set of eight polypeptides (Table 2A) that are indicative for first void urine was defined as the IPP, distinguishing first void from midstream urine samples. Three of these polypeptides (highlighted with asterisk in Table 2A) could also be found in seminal plasma reference samples, further supporting the initial hypothesis (results not shown). Utilizing an SVM model based on these eight polypeptides, 79/86 first void urine samples (47/51 PCa and 32/35 control) were found to be informative and 135/138 midstream samples to be “noninformative” using a threshold $F_{\text{informative}} = 0.3$ (Fig. 3A).

For definition of PCa specific biomarkers, the CE-MS data from patients with biopsy-proven PCa and men with benign biopsies were compared. The PCa samples of the training set appeared to be a nonhomogeneous cohort with respect to potentially PCa-indicative peptides. Initially defined biomarkers recognized PCa samples with only moderate sensitivity, and in general the same PCa samples were misclassified using different classification models (results not shown). To address this heterogeneity, an additional bootstrapping process for PCa specific biomarkers definition was performed: Only the subset of misclassified PCa samples of the training set was compared in a second experiment with the benign samples. Resulting marker candidates were combined with the initial markers to develop the final polypeptide signature of 12 polypeptides (Table 2B) indicative for PCa. After SVM model building, the discriminatory ability of the SVM-based classification factor $F$ to distinguish between patients with positive versus negative biopsies in the training set was assessed using ROC analysis (Fig. 3B). Using a threshold of $F_{\text{PCa}} = 0.0$, the training set was classified with sensitivity 80% (95% CI 67-90) and specificity 97% (95% CI 85-99) (Table 3).

The PCaP was evaluated in a blinded assessment of 264 initial urine samples. In a first step, all samples were examined employing the IPP (Fig. 1). Two hundred thirteen of these samples (80%) scored as informative (Table 3). The median classification factor $F$ of the informative panel in the group of biopsy positive men (1.67 ± 2.4, $N = 118$) did not differ ($p = 0.81$, Mann-Whitney test) from the median observed for the group of biopsy negative men (1.67 ± 2.8, $N = 95$). Subsequently, only informative samples were assessed with the PCaP using the classification factor threshold of $F_{\text{PCa}} = 0.0$. One hundred twenty-four samples scored as PCa and 89 as NED under these conditions (Fig. 1). After unblinding, 86/118 PCa samples (sensitivity 73% (95% CI 64-81)) and 57/95 NED samples (specificity 60% (95% CI 49-69)) were classified according to the obtained initial biopsy result (Figs. 4A and B). The median classification factor $F$ of the PCa panel in the group of biopsy-positive men (0.69 ± 1.2, $N = 118$) significantly differed ($p<0.0001$, Mann-Whitney test) from the median observed for biopsy-negative men (-0.29 ± 1.6, $N = 95$). Adjusting the observed ROC curve [47] according to the reported false negative rate of prostate biopsy (at least 30% [48-50]) results in improved proteomic test characteristics (75 and 68% sensitivity and specificity, respectively). A subclassification of the blinded set of samples according to T-PSA levels revealed: AUC = 0.76 for PSA<4, AUC = 0.78 for 4<PSA<10, and AUC = 0.65 for PS>10. The observed differences were statistically insignificant ($p = 0.179$ for PSA<4 compared to PSA>10 and $p = 0.102$ for 4<PSA<10 compared to PSA>10). In addition to sensitivity and specificity, positive and negative predictive values (PPV and NPV, respectively) for different prevalences were calculated (Fig. 4C).

In order to assess, if a combination of our proteomic classification result with available clinical and laboratory data results in additionally improved prediction of risks of positive biopsy, a
nomogram was established [51,52]. In multivariate analysis of prebiopsy risk factors (age, F-factor, percent F-PSA, T-PSA, and TRUS PV), PV and T-PSA were no significant independent predictors of a positive prostate biopsy (Table 4) in the set of 109 samples (60 PCa and 49 NED). Therefore, a nomogram was created using age (45-90 years), F-factor (-2 to 2), and F/T-PSA (30 to 0%) and classified the set with sensitivity of 90% and specificity of 61%. The predictive accuracy of the nomogram (AUC = 0.82, Fig. 4C) was found to be superior to that obtained for the proteomic F-factor (p = 0.008), percent free PSA (p = 0.009) or age (p = 0.004) alone. Compared to T-PSA (AUC = 0.57) that is used in routine screenings for PCa, the nomogram provides a highly significant (p<0.001) improvement. Compared to a reference nomogram consisting of percent free PSA and age alone (AUC = 0.77), the proteome pattern resulted in an improvement. Also comparison of PPV and NPV for a cutoff corresponding to 90% sensitivity in both nomograms and assuming a prevalence of 55% (as observed in the cohort investigated) showed a benefit, when including the proteome pattern. PPV increased from 68 to 74%, and NPV increased from 80 to 83%. However, the number of available samples was too small to demonstrate that the observed benefit of the nomogram including PCaP is of statistic significance (p = 0.139).

To identify polypeptides indicative of PCa, we have performed extensive MS/MS analysis (Table 2B). High-resolution MS/MS spectra could be obtained for a variety of polypeptides, with the majority of fragments identified derived from different types of collagens. The spectra of four candidates, a 1588.8, 1680.82, 2282.07, and 3409.72 Da peptides matched to entries of our classification models and are shown in Table 5. Sodium/potassium-transporting ATPase γ [2-18] (Fig. 5A), Collagen α-1 (III) [642-659] (Fig. 5B), Collagen α-1(I) [705-725], and Psoriasis susceptibility 1 candidate gene 2 protein [75-105] (also called SPR1) were identified. In addition, fragments of glioma tumor suppressor candidate region gene 1, hepatocellular carcinoma associated protein TB6, histone H2B, osteopontin, polymeric Ig receptor, transmembrane secretory component, prostatic acid phosphatase, Prostate-specific antigen, fibrinogen alpha chain precursor, and semenogelin 1 were also identified. Many of these peptides have existing connections to either the prostate and/or to multiple carcinomas and their subsequent disease states.

4 Discussion

The aim of our study was to establish and validate a proteome-based noninvasive approach for the detection of PCa with improved sensitivity and specificity compared to the currently available standards of DRE and serum markers (PSA and percent free PSA) for patients suspected of harboring PCa. This was motivated in part because the specificity of PSA in combination with DRE remains unsatisfactory in the range from 2 to 20 ng/mL, leading to almost 80% of all prostate biopsies being negative [53]. Several recent studies reported on biomarkers for PCa in serum [54-57]. So far, many of these studies are based on relatively small numbers of patients and the data still await validation utilizing samples collected in blinded prospective studies. Perhaps the most promising current marker for PCa appears to be early prostate cancer antigen (EPCA) [58-60].

To date, most urinary polypeptide markers are identified with immunochemical assays [61-66]. The first proteomic profiling experiments of urine to detect PCa were performed by Rehman et al. [67], who analyzed urine of six patients suffering from PCa and six BPH controls after prostate massage using 2-DE MS fingerprinting [67]. Urinary calgranulin B/MRP-14 was identified as a potential novel marker for PCa. M’Koma et al. [68] introduced a method for discriminating PCa, PIN3, and BPH by urine analysis using RP adsorption followed by MALDI-TOF MS. The authors were able to discriminate PCa and BPH with 67.4% sensitivity and 71.2% specificity, and PCa and HGPIN resulted in 81.0% sensitivity and 80.8% specificity. However, prospective validation of these results remains to be carried out.
In this work we defined PCa specific biomarkers and evaluated their diagnostic performance in a blinded prospective multicenter study of patients undergoing prostate biopsy. Our study also utilized extensive controls from patients with nonmalignant genitourinary diseases to establish quality control criteria for the samples and used different clinical centers to generalize the results as proposed by standards [27]. Several publications suggest polypeptides originating from prostatic fluid as potential biomarkers for PCa [44-46]. Since prostatic fluid is preferentially present in the first voided urine [31,32], we focused on this urine fraction as the assay fluid for our study. In order to cope with the problem of varying amounts of prostatic fluid in the urine samples, a polypeptide panel that should be indicative for the presence of sufficient prostatic fluid was established (the IPP). This panel was utilized to exclude noninformative samples containing very low or no prostatic fluid from data interpretation. These noninformative samples score as negative biopsy or healthy control even if they were collected from PCa patients, since the PCa specific biomarkers are missing. These findings show that quality control criteria are essential to avoid a large number of false negative classifications and ensure reproducible analysis when utilizing high resolution proteome analysis [19]. In our study, approximately 20% of the collected first void urine samples were excluded as noninformative (Table 3). Since the presence of an informative pattern may vary as a function of the prostatic secretions present in the urine, resampling of urine of non-informative patients (as indicated in Fig. 1) should be helpful in evaluating as many patients as possible with this approach, since IPP failure rate is often due to negligent sampling. Most patients, who were asked for a second sample and who were enlightened again on possible pitfalls, in fact supplied a sample that passed IPP criteria. This assumption remains to be further evaluated in larger studies.

For subsequent application of the PCa specific biomarker panel, only urine samples that were classified as informative were utilized. The low specificity of 60% (95% CI 49-70) in the 213 informative samples of the blinded test set reflects the complexity of the classification problem. These results may also reflect the fact that 30% or more of patients with negative biopsies are reported to harbor PCa at later biopsy [48-50]. Therefore, we will continue to follow patients with negative biopsies, and if they do develop PCa, we will refine the polypeptide panel accordingly. So far, four patients in the test set with negative first biopsy and a high proteomics-based probability of PCa underwent a second biopsy. In all four cases, the second biopsy detected the tumor.

For the clinical application of our model, a shift of the used classification threshold from 0.0 to -0.65 appears to be advisable to improve the sensitivity of our set-up from 73 to 88% (42% specificity). This adjustment ensures that the NPV, i.e., the portion of patients with negative diagnostic pattern that have in fact no PCa, is constantly above 80% for PCa prevalences<50% (Fig. 4C), giving highest possible diagnostic certainty for the decision, whether prostate biopsy may be protracted or even forborne in favor of watchful waiting. Adjusting the observed ROC curve [47] according to the false negative rate of prostate biopsy (>30%) results in the proteomic test having at least 89 and 51% sensitivity and specificity, respectively.

Our study is mainly composed of patients seen in university hospital settings, which tend to select for patients that have a more complex clinical course. This may account for the higher than usual number of patients with positive biopsies in this setting in both the training and test cohorts. To ensure that the generated SVM model generalizes to patients at other centers, we have included samples from international sites and private urological offices in the prospective arm of our study. The combination of the proteomic classification result with available clinical data suggests additionally improved prediction of risks of positive biopsy. The provided nomogram based on age, percent free PSA, and proteomic markers resulted in a sensitivity of 90% and specificity of 61%. If adjustments [47] are made to account for the false negative rate of prostate biopsy (at least 30% [48-50]) a sensitivity of 91% and a specificity of 69% can be
observed. However, the number of samples was too small to demonstrate that the observed benefit of the nomogram. Future randomized prospective studies are required to quantify the net gain of this approach.

Besides many peptides that have existing connections to multiple carcinomas, to date four of the defined biomarkers could be successfully sequenced using state-of-the-art MS/MS: sodium/potassium-transporting ATPase $\gamma$ [2-18] as a member of the informative panel (Tables 2A and 4), and Collagen $\alpha$-1 (III) [642-659], Collagen $\alpha$-1 (I) [699-725], and Psoriasis susceptibility 1 candidate gene 2 protein [75-105] (also called SPR1) as biomarkers of PCa (Tables 2B and 4). The main function of ATPase $\gamma$ is to maintain sodium and potassium homeostasis in cells. Interestingly, this molecule was found to be decreased in patients with positive biopsies compared to those with negative biopsies [69]. Collagen $\alpha$-1 (III) and Collagen $\alpha$-1 (I) are substrates of matrix metalloproteinases (MMP), a group of zinc finger endopeptidases with partially overlapping substrate specificity [70]. Regulation of MMP activity has been found for different cancers [71]. The collagen fragments found in this study were down-regulated in PCa patients compared to those with negative prostate biopsy (Table 2B). Expression of psoriasis susceptibility 1 candidate gene 2 protein [75-105], also known as small proline-rich protein 1, is detected in various diseases and cancers [72-74]. Down-regulation of SPR1, found in PCa samples as compared to NED samples (Table 2B), has been previously described for bronchogenic carcinoma cell lines measured by RT-PCR [75]. Interestingly, the molecular mass of marker protein 30566 (3425.68 Da, migration time of 31.3 min, Table 2B) matches well with the hypothesized mass of an oxidized form of the sequenced SPR1 fragment (ΔM = 116). As found in the cited RT-PCR study, protein 30566 was also found to be down-regulated in PCa samples. The majority of marker candidates in this study were found to be down-regulated in patients with PCa compared to patients with negative biopsies. On a conceptual level, this observation might be explained by the activation of yet unknown proteases by malign changes of the prostate. Subsequently, this activation might result in degradation of polypeptides that are typically observed in urine of patients with benign prostatic changes.

In summary, we have shown in a blinded multicentric prospective study that a panel of 12 urinary peptides present in initial voided urine, in combination with age and F/T-PSA serves as a predictor for the presence of PCa on biopsy and may give guidance to patients and clinicians whether additional evaluation by prostate (re-)biopsy should be considered.

Acknowledgments

This work was supported by NIH CA104106 to D. T. and Eurotransbio grant ETB-2006-016 to the Urosysteomics consortium (www.urosysteomics.com) to E. S. and H. M. The study sponsors had no role in study design, data collection, data analysis, data interpretation, or writing of the report. We are grateful to Axel Semjonow for supplying samples to this study and Peer Köster, Isabelle Butkay, Marco Schiemann, Charles Owens, Eric Blum, Harriet Lee, Francoise Flammang, Hilla Bürgel, and Beate Pepping-Schefers for assistance with sample handling, collection, or CE-MS measurements. J. J. C. and D. M. G. thank the University of Wisconsin-Madison, Thermo Fisher, and the National Institutes of Health (1R01GM080148 to J. J. C.) for financial support of this work. D. M. G. gratefully acknowledges support from an NIH predoctoral fellowship - the Biotechnology Training Program (NIH 5T32GM08349).

Abbreviations

DRE, digital rectal examination; F/T-PSA, percent free serum prostate specific antigen; IPP, “informative” polypeptide panel; NED, no evidence of disease; PCa, prostate cancer; PCaP, prostate cancer specific pattern; ROC, receiver operating characteristic curve; SVM, support vector machine; T-PSA, total serum prostate specific antigen; TRUS, transrectal ultrasound.
5 References


[7]. Ornstein DK, Rayford W, Fusaro VA, Conrads TP, et al. Serum proteomic profiling can discriminate prostate cancer from benign prostates in men with total prostate specific antigen levels between 2.5 and 15.0 ng/ml. J. Urol 2004;172:1302–1305. [PubMed: 15371828]


Figure 1.
Flow chart describing dual panel analysis used to evaluate urine samples from 264 patients undergoing prostate biopsy in the prospective test set.
* Using a classification factor F=0.0 (see Materials and Methods)
Figure 2. ROC evaluation of the nomogram predicting the probability of a positive prostate biopsy. Only those samples of the test set (Table 1A) for which information on F/T-PSA was available (in addition to T-PSA, patient age, and TRUS PV) were used for nomogram development (Table 4). The cohort included 109 informative urine samples (60 PCa and 49 NED). The predictive accuracy of the nomogram (Table 4, black, AUC = 0.82) was found to be improved ($p = 0.008$) compared to that obtained for the proteomic $F$-factor alone (gray, AUC = 0.70). Observed sensitivity and specificity of the nomogram (bold curve, threshold $p = 0.50$) is 90 and 61%, respectively. Dashed curve represents the adjusted ROC (AUC = 0.84) considering 30% false negative rate of prostate biopsy [48]. Adjusted sensitivity and specificity are 91 and 69%, respectively. Adjustment was made according to Punglia et al. [47].
Figure 3.
ROC curves (bold lines) of training sets used to establish the dual panel approach. (A) Using the informative polypeptides listed in Table 2A, 86 male first void urine samples (Table 1A) were distinguished from 138 male midstream urine samples from men without known PCa with high accuracy (AUC = 0.98). Seventy-nine of 86 first void urine samples (47/51 PCa and 32/35 NED) were found to be informative and 135/138 midstream samples to be noninformative using a threshold $F_{\text{informative}} = 0.3$. (B) The PCaP polypeptide panel from Table 2B was applied to 86 samples of the training set (51 PCa and 35 NED sample). A sensitivity of 80% and specificity 97% was found for classification factor $F_{\text{PCa}} = 0.0$. Dashed curves represent the 95% CI.
Figure 4.
Proteomics-based PCa diagnosis. (A) ROC curve of 213 informative samples of the test set using the proteomics dual panel approach (bold, AUC = 0.70). Observed sensitivity and specificity of the dual panel classification (threshold $F_{PCa} = 0.0$) is 73 and 60% respectively. Dashed curve represents the adjusted ROC (AUC = 0.77) considering 30% false negative rate of prostate biopsy [48]. Adjusted sensitivity and specificity are 74 and 68%, respectively. Adjustment was made according to Punglia et al. [47]. For the clinical application of the test, a shift of the used classification threshold from 0.0 to -0.65 appears to be advisable to improve sensitivity to 88% (42% specificity (unadjusted)) and 90% (51% specificity (adjusted)), respectively. (B) Box-and-whisker plots of classification factor $F$ obtained for classification of test set (Table 1A). The boxes depict the quartiles $Q_1$ and $Q_3$ of each distribution; the statistical medians are shown as horizontal lines in the boxes. The whiskers indicate 3/2 times the interquartile range of $Q_1$ and $Q_3$. (C) Positive (gray) and negative predictive values (black) are plotted against the prevalence of PCa (between 0 and 100%) for sensitivities and specificities observed for the IPP/PCaP classification in the blinded test set using either classification threshold $F_{PCa} = 0.00$ (73 and 60% sensitivity/specificity, bold lines) or using a shifted classification threshold $F_{PCa} = -0.65$ (88 and 42% sensitivity/specificity, dashed lines). The shift results in negative predictive values of constantly above 80%, i.e., the portion of
patients with negative diagnostic pattern that have in fact no PCa, for clinically expected PCa prevalences below 50% in cohorts of patients with elevated PSA levels.
Figure 5.
High-resolution MS/MS spectrum of (A) 1.589 kDa polypeptide indicative for PCa (Tables 2B and 5). Swiss-Prot database matching indicated this is a fragment of sodium/potassium transporting ATPase γ (2-18, *Homo sapiens*) with a calculated mass of 1588.70 Da and a sequence of TGLSMDGGGGSPKGDVDP. (B) Similar analysis for fragment of collagen alpha 1 type 3 chain [642-659] (*H. sapiens*) Mass: 1680.76 (Tables 2B and 5).
## Table 1A

Entire patient cohort uses

<table>
<thead>
<tr>
<th>Clinical condition</th>
<th>Patients (N)</th>
<th>Urine sample</th>
<th>Primary use</th>
<th>Secondary use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial urine samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training set</td>
<td>86</td>
<td>Initial 10 mL</td>
<td>Training set to develop informative panel</td>
<td>Training set to develop PCa panel</td>
</tr>
<tr>
<td>PCa</td>
<td>51</td>
<td>Initial 10 mL</td>
<td>Training set to develop informative panel</td>
<td>Training set to develop PCa panel</td>
</tr>
<tr>
<td>NED</td>
<td>35</td>
<td>Initial 10 mL</td>
<td>Training set to develop informative panel</td>
<td>Training set to develop PCa panel</td>
</tr>
<tr>
<td>Test set</td>
<td>264</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCa</td>
<td>118</td>
<td>Initial 10 mL</td>
<td>Test set to evaluate PCa panel</td>
<td>-</td>
</tr>
<tr>
<td>NED</td>
<td>95</td>
<td>Initial 10 mL</td>
<td>Test set to evaluate PCa panel</td>
<td>-</td>
</tr>
<tr>
<td>Midstream urine samples</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>138</td>
<td>Midstream</td>
<td>Training set to develop informative panel</td>
<td>-</td>
</tr>
<tr>
<td>Female</td>
<td>46</td>
<td>Midstream</td>
<td>Quality set to develop informative panel</td>
<td>-</td>
</tr>
</tbody>
</table>

PCa, prostate cancer on TRUS guided biopsy; NED, no evidence of PCa or HGPIN on TRUS guided biopsy.
<table>
<thead>
<tr>
<th></th>
<th>Training set (N = 86)</th>
<th>Test set (N = 264)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>PCa</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>65</td>
<td>11</td>
</tr>
<tr>
<td>T-PSA (ng/mL)</td>
<td>11.6</td>
<td>6.6</td>
</tr>
<tr>
<td>PV (mL)</td>
<td>48</td>
<td>32</td>
</tr>
<tr>
<td>Gleason sum</td>
<td>6.5</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>NED</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>63</td>
<td>9</td>
</tr>
<tr>
<td>T-PSA (ng/mL)(^a)</td>
<td>9.2</td>
<td>6.5</td>
</tr>
<tr>
<td>PV (mL)</td>
<td>54</td>
<td>22</td>
</tr>
</tbody>
</table>

PCa: prostate cancer on TRUS guided biopsy; NED: No evidence of PCa or HGPIN on TRUS guided biopsy; T-PSA: total serum prostate specific antigen

\(^a\) Measured by the "Access" Beckman-Coulter-Hybritech assay system.
Table 2A
The IPP: Eight polypeptide markers characteristic of initial voided urine derived from 86 samples in the training set and 138 male midstream urine samples described in Table 1A

<table>
<thead>
<tr>
<th>Polypeptide ID</th>
<th>Mass (Da)</th>
<th>CE-time (min)</th>
<th>Initial urine</th>
<th>Midstream urine (male)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Frequency</td>
<td>Median</td>
<td>IQR</td>
</tr>
<tr>
<td>10499</td>
<td>1588.8</td>
<td>30.2</td>
<td>90</td>
<td>462</td>
<td>516</td>
</tr>
<tr>
<td>11125*</td>
<td>1640.8</td>
<td>28.1</td>
<td>83</td>
<td>137</td>
<td>161</td>
</tr>
<tr>
<td>12569</td>
<td>1755.0</td>
<td>31.4</td>
<td>100</td>
<td>7943</td>
<td>8672</td>
</tr>
<tr>
<td>13273</td>
<td>1813.8</td>
<td>31.7</td>
<td>98</td>
<td>1841</td>
<td>1190</td>
</tr>
<tr>
<td>15104</td>
<td>1955.9</td>
<td>28.1</td>
<td>81</td>
<td>320</td>
<td>339</td>
</tr>
<tr>
<td>42855*</td>
<td>10324.3</td>
<td>23.0</td>
<td>86</td>
<td>1698</td>
<td>2383</td>
</tr>
<tr>
<td>42931*</td>
<td>10640.4</td>
<td>19.7</td>
<td>78</td>
<td>339</td>
<td>2393</td>
</tr>
<tr>
<td>42979*</td>
<td>10753.7</td>
<td>19.7</td>
<td>81</td>
<td>881</td>
<td>8953</td>
</tr>
</tbody>
</table>

*ID, polypeptide identifier annotated by the SQL database (ID) as described in Section 2. Polypeptides highlighted with asterisk could be verified in CE-MS profiled seminal plasma samples (results not shown); Mass, molecular weight (Da); CE-time, migration time (min) normalized to an array of 200 polypeptides frequently found in urine [33]; Frequency, percentage of samples in which the polypeptide was detected (%; between 0 = entirely absent and 100 = present in all samples) in initial urine samples (N = 86) compared to male midstream samples without known PCa (N = 138, Table 1); Median, median of normalized signal intensities; IQR, inter quartile range of normalized signal intensities.

* Primary or secondary mass (see Section 3).
Table 2B
The PCaP polypeptide panel: Polypeptide markers derived from training set of 51 PCa and 35 patients with negative biopsy (NED)

<table>
<thead>
<tr>
<th>Polypeptide ID</th>
<th>Mass (Da)</th>
<th>CE-time (min)</th>
<th>PCa</th>
<th>NED</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Median</td>
<td>IQR</td>
<td>Frequency</td>
<td>Median</td>
</tr>
<tr>
<td>10054</td>
<td>1552.56</td>
<td>37.2</td>
<td>96</td>
<td>100</td>
<td>0.032</td>
</tr>
<tr>
<td>10442</td>
<td>1584.58</td>
<td>37.7</td>
<td>78</td>
<td>83</td>
<td>0.049</td>
</tr>
<tr>
<td>11645</td>
<td>1680.82</td>
<td>30.0</td>
<td>73</td>
<td>97</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>12064</td>
<td>1714.62</td>
<td>37.9</td>
<td>57</td>
<td>77</td>
<td>0.038</td>
</tr>
<tr>
<td>12202</td>
<td>1725.66</td>
<td>38.4</td>
<td>100</td>
<td>100</td>
<td>0.001</td>
</tr>
<tr>
<td>14707</td>
<td>1925.90</td>
<td>23.3</td>
<td>71</td>
<td>60</td>
<td>0.050</td>
</tr>
<tr>
<td>19070</td>
<td>2282.07</td>
<td>34.0</td>
<td>100</td>
<td>97</td>
<td>0.013</td>
</tr>
<tr>
<td>27863</td>
<td>3092.00</td>
<td>29.7</td>
<td>96</td>
<td>100</td>
<td>0.020</td>
</tr>
<tr>
<td>30451</td>
<td>3409.72</td>
<td>32.2</td>
<td>82</td>
<td>94</td>
<td>0.028</td>
</tr>
<tr>
<td>30566</td>
<td>3425.68</td>
<td>31.3</td>
<td>100</td>
<td>10.0</td>
<td>0.031</td>
</tr>
<tr>
<td>32748</td>
<td>3765.54</td>
<td>20.2</td>
<td>63</td>
<td>40</td>
<td>0.011</td>
</tr>
<tr>
<td>43009</td>
<td>10770.20</td>
<td>19.6</td>
<td>55</td>
<td>71</td>
<td>0.010</td>
</tr>
</tbody>
</table>

ID, polypeptide identifier annotated by the SQL database (ID) as described in Section 2; Mass, molecular weight (Da); CE-time, migration time (min) normalized to an array of 200 polypeptides frequently found in urine [33]; PCa, prostate cancer on TRUS guided biopsy; NED, no evidence of PCa or HGPIN on TRUS guided biopsy; Frequency, percentage of samples in which the polypeptide was detected (%, between 0 = entirely absent and 100 = present in all samples) in PCa urine samples (N = 51) compared to NED urine samples (N = 35, Table 1); Median, median of normalized signal intensities; IQR, interquartile range of normalized signal intensities.

Marked *p*-values according to Wilcoxon test and Benjamini and Hochberg (BH) for multiple testing corrections are related to a PCa sub-population (see Section 2).

* Primary or secondary mass (see Section 3).
Table 3
Classification of initial urine samples (Table 1) using IPP (Table 2A) and PCa panel (Table 2B) as shown in Fig. 1

<table>
<thead>
<tr>
<th>Clinical condition</th>
<th>Total samples (N)</th>
<th>Informative samples (N)</th>
<th>False predicted biopsy result (N)</th>
<th>Correctly predicted biopsy result (N)</th>
<th>Correct classification (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCa</td>
<td>51</td>
<td>47</td>
<td>10</td>
<td>41</td>
<td>80% (67-90)</td>
</tr>
<tr>
<td>NED</td>
<td>35</td>
<td>32</td>
<td>1</td>
<td>34</td>
<td>97% (85-99)</td>
</tr>
<tr>
<td>Test set</td>
<td>264</td>
<td>213</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCa</td>
<td>142</td>
<td>118</td>
<td>32</td>
<td>86</td>
<td>73% (64-81)</td>
</tr>
<tr>
<td>NED</td>
<td>122</td>
<td>95</td>
<td>38</td>
<td>57</td>
<td>60% (49-70)</td>
</tr>
</tbody>
</table>

PCa: prostate cancer on TRUS guided biopsy; NED: No evidence of PCa or HGPIN on TRUS guided biopsy; CI: Confidence Interval

*a* Using a classification factor $F = 0.0$ (see Section 2).
Table 4
Development of a nomogram using T-PSA, percent F-PSA, PV, and proteomics classification factor F of 109 patients

<table>
<thead>
<tr>
<th>Variable</th>
<th>p-Value</th>
<th>Coefficient</th>
<th>Odds ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>&lt;0.01</td>
<td>0.119</td>
<td>1.13</td>
<td>1.04-1.22</td>
</tr>
<tr>
<td>F-factor</td>
<td>&lt;0.01</td>
<td>0.931</td>
<td>2.54</td>
<td>1.40-4.59</td>
</tr>
<tr>
<td>F/T-PSA</td>
<td>&lt;0.01</td>
<td>-0.183</td>
<td>0.83</td>
<td>0.76-0.92</td>
</tr>
<tr>
<td>T-PSA</td>
<td>0.71</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PV</td>
<td>0.90</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Constant</td>
<td>-5.819</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Significance level <0.001

F-factor, euclidian distance of data points to the separating hyper plane; PV, TRUS prostate volume (mL).
Table 5

<table>
<thead>
<tr>
<th>Polypeptide ID</th>
<th>Experimental mass (Da)</th>
<th>Migration time (min)</th>
<th>Sequence</th>
<th>Name</th>
<th>Calculated mass (Da)</th>
<th>Mass deviation (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCaP 18645</td>
<td>1680.821</td>
<td>30.0</td>
<td>GLPhGTGPPhGENGKPhGEPh</td>
<td>Collagen alpha-1 (III) chain [642-659]</td>
<td>1680.759</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>19070</td>
<td>34.0</td>
<td>ANGA/P6GNDKGDAGA/P6GAPhGAPhGSQGAPhG</td>
<td>Collagen alpha-1 (I) chain [699-725]</td>
<td>2281.979</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>30451</td>
<td>32.2</td>
<td>DLPETGVWPPEPPRTDPPQPPRDWPAGP</td>
<td>Psoriasis susceptibility 1 candidate gene 2 protein [75-105] (also called SPR1)</td>
<td>3409.637</td>
<td>23</td>
</tr>
<tr>
<td>IPP 10499</td>
<td>1588.772</td>
<td>30.2</td>
<td>TGLSMGD/GGSKGDVDVP</td>
<td>Sodium/potassium-transporting ATPase gamma chain [2-18]</td>
<td>1588.704</td>
<td>43</td>
</tr>
</tbody>
</table>