[CANCER RESEARCH 62, 3609–3614, July 1, 2002]

Advances in Brief

Serum Protein Fingerprinting Coupled with a Pattern-matching Algorithm Distinguishes Prostate Cancer from Benign Prostate Hyperplasia and Healthy Men[1]

Bao-Ling Adam,[2] Yinsheng Qu,[2] John W. Davis, Michael D. Ward, Mary Ann Clements, Lisa H. Cazares, O. John Semmes, Paul F. Schellhammer, Yutaka Yasui, Ziding Feng, and George L. Wright, Jr.[2,3]

Departments of Microbiology and Molecular Cell Biology [B-L. A., M. D. W., M. A. C., L. H. C., O. J. S., G. L. W.] and Urology [P. F. S., G. L. W.] and Virginia Prostate Center [B-L. A., J. W. D., M. D. W., M. A. C., L. H. C., O. J. S., P. F. S., G. L. W.], Eastern Virginia Medical School, Norfolk, Virginia, 23501; Sentara Cancer Institute, Norfolk, Virginia 23501 [B-L. A., J. W. D., M. D. W., M. A. C., L. H. C., O. J. S., P. F. S., G. L. W.]; Department of Urology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030 [J. W. D.]; and Fred Hutchinson Cancer Research Center, Seattle, Washington 98109 [Y. Q., Y. Y., Z. F.]

Abstract

The prostate-specific antigen test has been a major factor in increasing awareness and better patient management of prostate cancer (PCA), but its lack of specificity limits its use in diagnosis and makes for poor early detection of PCA. The objective of our studies is to identify better biomarkers for early detection of PCA using protein profiling technologies that can simultaneously resolve and analyze multiple proteins. Evaluating multiple proteins will be essential to establishing signature proteomic patterns that distinguish cancer from noncancer as well as identify all genetic subtypes of the cancer and their biological activity. In this study, we used a protein biochip surface enhanced laser desorption/ionization mass spectrometry approach coupled with an artificial intelligence learning algorithm to differentiate PCA from noncancer cohorts. Surface enhanced laser desorption/ionization mass spectrometry protein profiles of serum from 167 PCA patients, 77 patients with benign prostate hyperplasia, and 82 age-matched unaffected healthy men were used to train and develop a decision tree classification algorithm that used a nine-protein mass pattern that correctly classified 96% of the samples. A blinded test set, separated from the training set by a stratified random sampling before the analysis, was used to determine the sensitivity and specificity of the classification system. A sensitivity of 83%, a specificity of 97%, and a positive predictive value of 96% for the study population and 91% for the general population were obtained when comparing the PCA *versus* noncancer (benign prostate hyperplasia/healthy men) groups. This high-throughput proteomic classification system will provide a highly accurate and innovative approach for the early detection/diagnosis of PCA.

Introduction

The number of PCA[4] cases has tripled during the past decade due to the widespread use of serum PSA testing and DRE (1). Although these efforts have allowed for increased identification of individuals with cancer, overall "early" detection or determination of aggressive cancers is needed. PSA is currently the best overall serum marker for PCA in clinical use. Nevertheless, the PSA test lacks specificity (2, 3), limiting its use as an early detection biomarker, and its relation to biological activity has been questioned (4). It is important that additional diagnostic biomarkers be identified to reduce PCA mortality. However, because of the robust molecular and cellular heterogeneity of PCA, it is likely that a combination or a panel of biomarkers will be required to improve the early detection of PCA.

The study of the cell's proteome presents a new horizon for biomarker discovery. Two-dimensional PAGE has been the classical approach to explore the proteome for separation and detection of differences in protein expression (5, 6). Advances in two-dimensional gel electrophoresis technology coupled with robotics and software programs for identifying potential protein alterations have improved this proteomic system. Nevertheless, two-dimensional gel electrophoresis is still cumbersome, labor intensive, suffers reproducibility problems, and is not readily transformed into a clinical assay. Advances have also been made in mass spectrometry to achieve high-throughput separation and analysis of proteins (7–9). One of the recent advances is the ProteinChip system manufactured by Ciphergen Biosystems, Inc. (Fremont, CA). This system uses SELDI time-of-flight mass spectrometry to detect proteins affinity-bound to a protein chip array (10, 11). This system is a novel, extremely sensitive, and rapid method to analyze complex mixtures of proteins and peptides. Initial studies from our laboratory established the potential of SELDI for discovery and profiling of prostate and bladder cancer biomarkers in body fluids and cell lysates (12, 13).

The objective of this study was to determine whether SELDI protein profiling of serum coupled with an artificial intelligence data analysis algorithm could effectively differentiate PCA from BPH and unaffected HM. Using a standardized test set, we demonstrate proof of principle that our SELDI protein profiling approach can accurately discriminate PCA from patients with BPH and men of the same age who do not have prostate disease. Our results form the basis for initiating further evaluation and validation to assess the potential of this SELDI proteomic classification system for the early detection and diagnosis of PCA, and further study is warranted to establish profiles that identify the clinically important lethal cancers.

Materials and Methods

Serum Samples. Serum samples were obtained from the Virginia Prostate Center Tissue and Body Fluid Bank. The serum procurement, data management, and blood collection protocols were approved by the Eastern Virginia Medical School Institutional Review Board. Blood samples from patients diagnosed with either PCA or BPH were procured from the Department of Urology, Eastern Virginia Medical School, and the HM cohort was obtained from free screening clinics open to the general public. Only pretreatment samples obtained at the time of diagnosis of PCA or BPH were used for this study. After obtaining informed consent from the patient, the sample was

Received 3/29/02; accepted 5/15/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported by the National Cancer Institute Early Detection Research Network (http://edrn.nci.nih.gov; Grant CA85067) and the Virginia Prostate Center (http://www.evms.edu/vpc/).

[2] B-L. A., Y. Q., and G. L. W. contributed equally to this work.

[3] To whom requests for reprints should be addressed, at Department of Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, 700 West Olney Road, Norfolk, VA 23501. Phone: (757) 446-5662; Fax: (757) 624-2255; E-mail: wrightgl@evms.edu.

[4] The abbreviations used are: PCA, prostate cancer; PSA, prostate-specific antigen; SELDI, surface enhanced laser desorption/ionization mass spectrometry; BPH, benign prostate hyperplasia; HM, healthy men; EAM, energy-adsorbing molecule; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; DI, deionized. DRE, digital rectal examination.

collected into a 10-cc Serum Separator Vacutainer Tube and centrifuged 30 min later at 375,000 rpm for 5 min. The serum was distributed into 500-μl aliquots and stored frozen at −80°C. A quality control sample was prepared by pooling an equal amount of serum from each specimen of the age-matched HM group and storing 100-μl aliquots at −80°C. The quality control serum sample was used to determine reproducibility and as a control protein profile for each SELDI experiment.

Patient and Donor Cohorts. Specimens from four groups of patients were used in this study: (*a*) 97 age-matched HM (control); (*b*) 92 patients with BPH; (*c*) 99 patients diagnosed with organ-confined PCA (T_1/T_2); and (*d*) 98 patients diagnosed with non-organ-confined PCA (T_3/T_4). A donor was selected for the HM group if he had a normal DRE, a PSA < 4.0 ng/ml, and no evidence of prostatic disease. The HM group consisted of 48 Caucasian and 48 African-American males ranging in age from 51–70 years (mean age, 60 years). There were 33 Caucasians, 2 African Americans, and 57 men of unknown race in the BPH patient group, ranging in age from 48–86 years (mean age, 67 years). The BPH patients were selected if they had PSA values between 4 and 10 ng/ml, low PSA velocities (*i.e.*, PSA velocity <0.7 ng/ml/year), and multiple negative biopsies. The number of biopsies was two (73 cases), three (13 cases), and four (6 cases). The organ-confined PCA group (T_1/T_2) consisted of 76 Caucasians, 20 African Americans, 1 Asian, and 2 men of unknown race with ages ranging from 50–89 years (mean age, 71 years). For the non-organ-confined PCA group (T_3/T_4), there were 80 Caucasians, 16 African Americans, and 2 men of unknown race, ranging in age from 44–87 years (mean age, 69 years). The range and mean PSA values for the groups were as follows: a 0.15–3.83 ng/ml (1.32 ng/ml) for the HM group [86 members of this group had a PSA < 2.5 ng/ml (the latter were considered to be a low-risk group)]; (*b*) 0.0–10.91 ng/ml (4.60 ng/ml) for the BPH group; (*c*) 0.0–95.16 ng/ml (10.10 ng/ml) for the organ-confined PCA (T_1/T_2) group; and (*d*) 0.0–8752 ng/ml (206.93 ng/ml) for the non-organ-confined PCA (T_3/T_4) group.

SELDI Protein Profiling. Various chip chemistries (hydrophobic, ionic, cationic, and metal binding) were initially evaluated to determine which affinity chemistry provided the best serum profiles in terms of number and resolution of proteins. The IMAC-Cu metal binding chip was observed to give the best results. IMAC-3 chips (Ciphergen Biosystems, Inc.) were coated with 20 μl of 100 mM $CuSO_4$ on each array, placed on a TOMY Micro Tube Mixer (MT-360; Tomy Seiko Co., Ltd.), and agitated for 5 min. The chips were rinsed 10 times with DI water, and 20 μl of 100 mM sodium acetate were added to each array and shaken for 5 min to remove the unbound copper. The chips were rinsed again with DI water (10 times) and put into a bioprocessor (Ciphergen Biosystems, Inc.), which is a device that holds 12 chips and allows application of larger volumes of serum to each chip array. The bioprocessor was washed and shaken on a platform shaker at a speed of 250 rpm for 5 min with 200 μl of PBS in each well. This was repeated twice more, and each time the PBS buffer was discarded by inverting the bioprocessor on a paper towel. Serum samples for SELDI analysis were prepared by vortexing 20 μl of serum with 30 μl of 8 M urea/1% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid in PBS in a 1.5-ml microfuge tube at 4°C for 10 min. One hundred μl of 1 M urea with 0.125% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonic acid were added to the serum/urea mixture and vortexed briefly. PBS was added to make a 1:5 dilution and placed on ice until applied to a protein chip array. Fifty μl of the diluted serum/urea mixture were applied to each well, and the bioprocessor was sealed and shaken on a platform shaker at a speed of 250 rpm for 30 min. The serum/urea mixture was discarded, and the PBS washing step was repeated three times. The chips were removed from the bioprocessor, washed 10 times with DI water, air dried, and stored in the dark at room temperature until subjected to SELDI analysis. Before SELDI analysis, 0.5 μl of a saturated solution of the EAM sinapinic acid in 50% (v/v) acetonitrile, 0.5% trifluoroacetic acid was applied onto each chip array twice, letting the array surface air dry between each sinapinic acid application. Chips were placed in the Protein Biological System II mass spectrometer reader (Ciphergen Biosystems, Inc.), and time-of-flight spectra were generated by averaging 192 laser shots collected in the positive mode at laser intensity 220, detector sensitivity 7, and a focus lag time of 900 ns. Mass accuracy was calibrated externally using the All-in-1 peptide molecular mass standard (Ciphergen Biosystems, Inc.).

Data Analysis. The data analysis process used in this study involved three stages: (*a*) peak detection and alignment; (*b*) selection of peaks with the highest discriminatory power; and (*c*) data analysis using a decision tree algorithm. A stratified random sampling with four strata [PCA (T_1/T_2), PCA (T_3/T_4), BPH, and HM] was used to separate the entire data set into training and test data sets before the analysis. The training data set consisted of SELDI spectra from 167 PCA, 77 BPH, and 82 normal serum samples. The validity and accuracy of the classification algorithm were then challenged with a blinded test data set consisting of 30 PCA, 15 BPH, and 15 normal samples.

Peak Detection. Peak detection was performed using Ciphergen SELDI software versions 3.0 β and 3.0.[5] The mass range from 2,000–40,000 Da was selected for analysis because this range contained the majority of the resolved protein/peptides. The molecular masses from 0–2,000 Da were eliminated from analysis because this area contains adducts and artifacts of the EAM and possibly other chemical contaminants. Peak detection involved (*a*) baseline subtraction, (*b*) mass accuracy calibration, and (*c*) automatic peak detection. The software program calculates noise, peak area, and filter based on the criteria selected by the operator for data analysis. The settings used for this study were as follows: (*a*) fitting window width = 100 data points; (*b*) average noise = 10 points; (*c*) peak area calculated using the slope-based method; (*d*) low minimum valley depth = 10 times noise; (*e*) high minimum valley depth = 0.5 times noise; (*f*) low and high sensitivity of peak height = 10 and 2 times noise, respectively; (*g*) auto peak detection slider = 8 for mass range 2–4 kDa, 11 for mass range 4–8 kDa, and 8 for mass range 8–40 kDa. An average of 81 peaks was detected in each spectrum.

Peak Alignment. All of the labeled peaks from 772 spectra were exported from SELDI to an Excel spreadsheet. A PeakMiner algorithm,[6] developed in-house, was used to align peaks and perform statistical analysis. Peaks were first sorted by mass, and a mass error value was calculated for each peak. The mass error score, the measurement of mass difference between peak *X* and peak *X* + 1, is calculated for each peak using (*Mpx* − *Mpx* +1)/*Mpx*, where *Mpx* is the mass value of peak *X*. For example, if the mass error score was <0.18%, peak *X* and peak *X* + 1 would align into one peak, representing the same protein in each sample. If the mass error was >0.18%, then peak *X* and peak *X* + 1, would be considered two distinct peaks. This is an iterative process throughout all of the labeled peaks.

Feature Selection. The power of each peak in discriminating normal *versus* PCA, normal *versus* BPH, and BPH *versus* PCA was determined by estimating the AUC, which ranges from 0.5 (no discriminating power) to 1.0 (complete separation).

Decision Tree Classification. Construction of the decision tree classification algorithm was performed as described by Breiman *et al.* (14) with modifications,[7] using a training data set consisting of 326 samples (82 normal, 77 BPH, and 167 PCA samples). Classification trees split up a data set into two bins or nodes, using one rule at a time in the form of a question. The splitting decision is defined by presence or absence and the intensity levels of one peak. For example, the answer to "Does mass A have an intensity less than or equal to X" splits the data set into two nodes, a left node for yes and a right node for no. This splitting process continues until terminal nodes or leaves are produced or further splitting has no gain. Classification of terminal nodes is determined by the group ("class") of samples (*i.e.*, PCA, BPH, or HM) representing the majority of samples in that node. A "cost" function is calculated that reflects the heterogeneity of each node: $-\log L = -\Sigma nj \log(pj)$ where L is the likelihood of the multinomial distribution, nj is the number of samples in class j, and pj is the probability of class j. Peaks selected by this process to form the splitting rules are the ones that achieve the maximum reduction of cost in the two descendant nodes.

Statistical Analyses. The AUC was computed to identify the peaks with the highest potential to discriminate the three groups, based on the probability that the test result from a diseased individual is more indicative of disease than that from a nondiseased individual (15). A Bayesian approach was used to calculate the expected probabilities of each class in each terminal node (16), and their 95% confidence intervals were calculated using the posterior Dirichlet distribution (16). The 95% confidence intervals were calculated by generating and sorting 4000 samples for the posterior Dirichlet distribution, and the 100th and 3900th sample were considered as the lower and upper bounds of the 95% confidence intervals, respectively. Specificity was calculated as the ratio of the number of nondisease samples correctly classified to

[5] Internet address: www.chiphergen.com.
[6] Internet address: www.evms.edu/vpc/seldi.
[7] Internet address: http://140.107.129.65/stat_methods.htm.

the total number of nondisease samples. Sensitivity was calculated at the ratio of the number of correctly classified diseased samples to the total number of diseased samples. The PPV for the study population was calculated by dividing the number of true PCA positives by the sum of the number of true PCA positives plus the number of false PCA positives. The NPV for the study population was calculated by dividing the number of true negative nondisease samples (BPH/HM) by the sum of the number of false negative plus the number of true negative nondisease samples (BPH/HM). The PPV and NPV for PCA *versus* noncancer (BPH/HM) in the general population were calculated as follows: PPV (for population) = sensitivity * rho/[sensitivity * rho + (1 − specificity) * (1 − rho)]; and NPV (for population) = specificity * (1 − rho/[specificity) * (1 − rho) + (1 − sensitivity) * rho], where rho is prevalence in the population.

Results

Data Analysis. Peak detection using the SELDI software program detected 63,157 peaks in the 2–40-kDa mass range after analysis of 772 spectra (386 spectra in duplicate, with approximately 81 peaks/spectrum). Of these, 779 peaks were identified after the clustering and peak alignment process. The AUC was calculated for each of the 779 peaks. No single peak was identified that had an AUC of 1.0, indicating that there was not a peak detected that alone could completely separate two groups (*i.e.,* HM *versus* PCA, HM *versus* BPH, or BPH *versus* PCA) or three groups (PCA *versus* BPH *versus* HM). Of the 779 peaks, 124 had an AUC ≥ 0.62. Those with an AUC < 0.62 were considered irrelevant for classification. These 124 peaks identified in the training set were then used to construct the decision tree classification algorithm. Fig. 1 is a flow diagram that summarizes the process from peak detection to sample classification. The classification algorithm used nine masses between 4 and 10 kDa (4475, 5074, 5382, 7024, 7820, 8141, 9149, 9507, and 9656 Da) to generate 10 terminal nodes (L1−L10; Fig. 2*A*). Once the algorithm identifies the most discriminatory peaks, the classification rule is quite simple. For example, if an unknown sample has no peak at mass 7819.75 ("root" node) but has a peak at mass 7024.02, then the sample is placed in terminal node L1 and classified as PCA. If the sample is placed in L2, it will be assigned to BPH. Another example of this splitting process is shown in Fig. 2*B*, in which four masses between 5 and 10 kDa are used to assign 46 of the 167 PCA samples to terminal node L7. Based

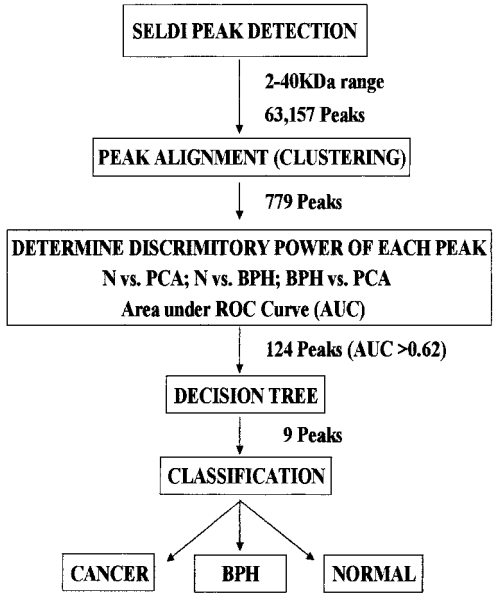

Fig. 1. Flow diagram showing the processes involved in development of the classification tree analyses program. *N* and *Normal*, unaffected HM.

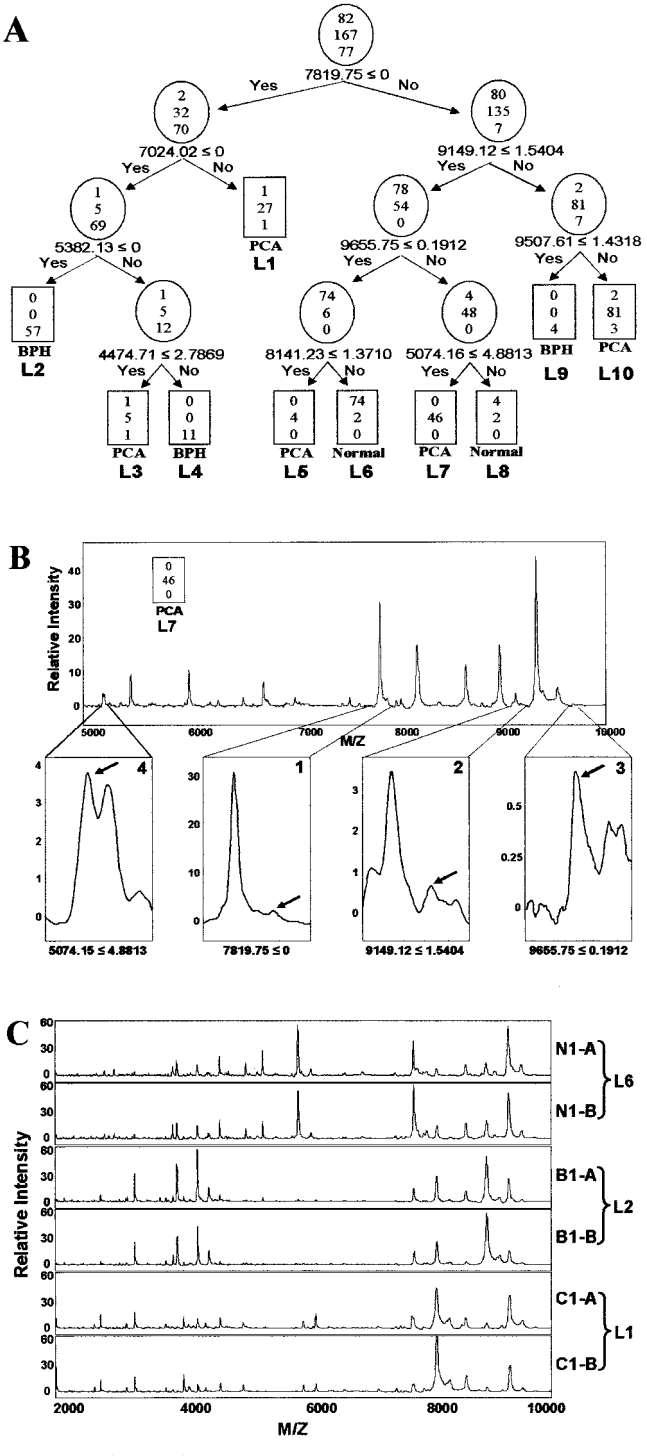

Fig. 2. Classification of the prostate disease and nondisease samples in the training data set. *A,* diagram of decision tree analyses. The root node (*top*) and descendant nodes are shown as *ovals*, and the terminal nodes (*L1−L10*) are shown as *rectangles*. The numbers in each node represent the classes [*top number*, number of HM (normal control) samples; *middle number*, number of PCA samples; *bottom number*, number of BPH 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 7819.75 kDa, and the intensity is ≤0. *B,* representative example of a SELDI spectrum showing the combination of four peak masses required to correctly classify the sample as PCA in the L7 terminal node. The *arrows* in the magnified panels identify the protein peaks used in the classifier, and the numbers *1–4* in the *top right corner* indicate the order the decision tree takes in assigning the sample to the L7 terminal node. The *first number* under each panel is the mass, and the *second number* is the peak intensity. *C,* example of the reproducibility of the SELDI and decision tree classification analyses. Serum samples randomly selected and repeated 18 months (*B*) after the initial SELDI analysis (*A*) showed similar spectra and were correctly classified to the appropriate terminal node by the decision tree algorithm; in this example either terminal node L1, L2, or L6. *N1,* sample from a healthy male donor; *B1,* sample from a patient with BPH; *C1,* sample from a patient with PCA.

Table 1 *Expected probabilities and the 95% confidence levels for each of the classes assigned to the 10 terminal nodes*

Node	Class	Observation	Probability	95% Confidence level (low, high)
L1	HM	1	0.0625	0.0081, 0.1693
	PCA	27	0.8750	0.7423, 0.9630
	BPH	1	0.0625	0.0087, 0.1698
L2	HM	0	0.0167	0.0005, 0.0584
	PCA	0	0.0167	0.0004, 0.0628
	BPH	57	0.9667	0.9072, 0.9952
L3	HM	1	0.2000	0.0247, 0.4793
	PCA	5	0.6000	0.3027, 0.8592
	BPH	1	0.2000	0.0248, 0.4753
L4	HM	0	0.0714	0.0018, 0.2509
	PCA	0	0.0714	0.0019, 0.2579
	BPH	11	0.8571	0.6311, 0.9823
L5	HM	0	0.1429	0.0040, 0.4725
	PCA	4	0.7143	0.3557, 0.9567
	BPH	0	0.1429	0.0040, 0.4504
L6	HM	74	0.9494	0.8950, 0.9858
	PCA	2	0.0380	0.0082, 0.0879
	BPH	0	0.0127	0.0003, 0.0459
L7	HM	0	0.0204	0.0005, 0.0738
	PCA	46	0.9592	0.8893, 0.9951
	BPH	0	0.0204	0.0005, 0.0726
L8	HM	4	0.5556	0.2458, 0.8416
	PCA	2	0.3333	0.0836, 0.6544
	BPH	0	0.1111	0.0032, 0.3732
L9	HM	0	0.1429	0.0034, 0.4560
	PCA	0	0.1429	0.0037, 0.4830
	BPH	4	0.7143	0.3354, 0.9566
L10	HM	2	0.0337	0.0068, 0.0784
	PCA	81	0.9213	0.8595, 0.9674
	BPH	3	0.0449	0.0123, 0.0964

on the stochastic nature of reality, misclassification of a new sample cannot be ruled out even for a pure node that contains only one sample type, for example, L2, which contains only BPH samples. To obtain an idea of whether an unknown sample would be correctly classified or misclassified, the expected probability and 95% confidence level was calculated for each class in the 10 terminal nodes (Table 1). The expected probabilities for HM and PCA samples to be misclassified in L2, for example, are 1.67%. Although not zero, the likelihood of HM or PCA samples being assigned to this node is extremely low; whereas BPH has a 96.67% chance of being correctly classified to L2 (with the 95% confidence interval between 90.72% and 99.52%). The probability of incorrect assignment of samples increases in nodes that contain few majority samples or when only a few samples are assigned to the node, as, for example, terminal nodes L3, L5, and L9 (Fig. 2*A*).

A summation of the classification results from the 10 terminal nodes is presented for the training and test sets in Table 2. The classification algorithm correctly predicted 93.51–97.59% of the samples for each of the three groups in the training set (Table 2A), for an overall correct classification of 96%. The algorithm correctly predicted 90% (54 of 60) of the test samples, with all 15 samples from HM, 93% (14 of 15) of the BPH samples, and 83% (25 of 30) of the PCA samples being correctly classified (Table 2B). Three of the misclassified HM cases in the training set had PSA values < 2.5 ng/ml (*i.e.*, 0.15, 0.76, and 1.52 ng/ml), considered a low-risk group, and the fourth case had a PSA of 3.02 ng/ml (*i.e.*, high-risk group). Therefore, no correlation for the misclassification of four of the HM cases with PSA levels could be made.

The sensitivity and specificity of the classification system for differentiation of disease from the nondisease groups are presented in Table 2C. When comparing PCA *versus* noncancer (BPH/HM), the sensitivity was 83% (25 of 30), and the specificity was 97% (29 of 30). A sensitivity of 83% was also obtained when comparing PCA *versus* HM (25 of 30) or PCA *versus* BPH (25 of 30), whereas the specificity was 100% (15 of 15) for PCA *versus* HM and 93% (14 of 15) for PCA *versus* BPH. The PPV and NPV for the study population were 96.15% and 96.67%, respectively. When considering an estimated 30% prevalence of PCA in the general population of men age 50 years or older (17), the PPV is 91.15%, and the NPV is 93.12%.

Reproducibility. The reproducibility of SELDI spectra, *i.e.*, mass location and intensity from array to array on a single chip (intra-assay) and between chips (interassay), was determined using the pooled normal serum quality control sample. Seven proteins in the range of 3,000–10,000 Da observed on spectra randomly selected over the course of the study were used to calculate the coefficient of variance. The intra-assay and interassay coefficient of variance for peak location was 0.05%, and the intra-assay and interassay coefficient of variance for normalized intensity (peak height or relative concentration) was 15% and 20%, respectively (data not shown). Masses that were within 0.18% mass accuracy between spectra were considered to be the same. Most important was the observation that randomly selected samples, blinded to the person performing SELDI and rerun

Table 2 *Decision tree classification of the prostate training and test sets*

Sample	Normal	BPH	PCA	Misclassified rate
A. Training set				
HM ($N = 82$)	78 (95.12%)	0 (0.00%)	4 (4.88%)	4 (4.88%)
BPH ($N = 77$)	0 (0.00%)	72 (93.51%)	5 (6.49%)	5 (6.49%)
PCA stage T_1, T_2 ($N = 84$)	2 (2.38%)	0 (0.00%)	82 (97.61%)	2 (2.38%)
PCA, stage T_3, T_4 ($N = 83$)	2 (2.40%)	0 (0.00%)	81 (97.59%)	2 (2.41%)
Total no. of samples ($N = 326$)				13 (3.99%)
Sample	Normal	BPH	PCA	Misclassified rate
B. Test set				
HM ($N = 15$)	15 (100.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
BPH ($N = 15$)	0 (0.00%)	14 (93.33%)	1 (6.67%)	1 (6.67%)
PCA, stage T_1, T_2 ($N = 15$)	3 (6.67%)	0 (0.00%)	12 (80.00%)	3 (20.00%)
PCA, stage T_3, T_4 ($N = 15$)	1 (6.67%)	1 (6.67%)	13 (86.67%)	2 (13.33%)
Total no. of samples ($N = 60$)				6 (10.00%)

	Percent positive (no. positive/no. tested)					
Disease/nondisease	PCA/HM	PCA/BPH	PCA/(BPH/HM)	BPH/HM	BPH/T1,T2	Ti,T2/T3,T4
C. Differentiation of prostate disease from nondisease in the blinded test set						
Sensitivity	83 (25/30)	83 (25/30)	83 (25/30)	93 (14/15)	93 (14/15)	80 (12/15)
Specificity	100 (15/15)	93 (14/15)	97 (29/30)	100 (15/15)	80 (12/15)	87 (13/15)

months or even a year later, were correctly classified by the decision tree classification algorithm (Fig. 2*C*).

Discussion

The current standard screening approach for PCA is a serum test for PSA, and if the test is positive, biopsies are obtained from each lobe of the prostate. Although the PSA test has a sensitivity of >90%, its specificity is only 25%. This low specificity results in subjecting men to biopsies of the prostate as well as considerable anxiety when they do not have PCA detectable by biopsy. With the SELDI profiling classification approach, an overall sensitivity of 83%, a specificity of 97%, and a PPV of 96% were obtained in differentiating PCA from BPH and age-matched unaffected HM. Provided that this SELDI profiling classification system can be validated using a larger and more clinically diverse study set, this approach would have immediate and substantial benefit in reducing the number of unnecessary biopsies.

Our successful development of a diagnostic system that achieved a high PPV (96%) for the blinded test set is based on using a large, carefully chosen training set of randomly selected samples. All specimens were closely age matched. Serum samples from unaffected HM, identified as men with a negative DRE and PSA <4.0 ng/ml, were obtained from the general population during free prostate screening clinics. Nevertheless, selecting a cancer-free control population for studies described herein is difficult. It is unusual for a man with a normal PSA and normal DRE to undergo a prostate biopsy to be certain that the controls are truly negative. About the best that can be done is to select healthy controls that have potentially the lowest risk for PCA. For this study, 86 of the 96 HM cases had PSA values <2.5 ng/ml, which is considered a lower-risk group. The majority of the BPH patients had 4–10 ng/ml PSA and multiple negative biopsies, and the PCA patients had cancers ranging from small volume localized disease to local and distant metastatic disease and PSA values varying from 0 to >8000 ng/ml. Another important factor in the construction of a successful classification system was using an algorithm that could filter out the "noise" that is characteristic of mass spectrometry instruments, the spurious signals created by the EAM and chemical contaminates introduced in the assay, and the natural random daily fluctuations and sample-to-sample variability. This "normalization" process is critical in distinguishing peaks due to artifacts from the true peptide/protein peaks. It becomes even more important when considering that most all of the protein alterations between the cancer and noncancer cohorts are based on the overexpression or underexpression of proteins and not solely on their presence or absence. We believe that accurate and reproducible feature selection or peak "picking" algorithms with normalization functions is the most critical first step in developing a successful classification algorithm for the SELDI profiling data.

It was encouraging that the three study cohorts could be separated based on the overexpression or underexpression of nine peptide/protein masses. However, it was not surprising that multiple biomarkers would be required to effectively deal with the problem of tumor microheterogeneity that has plagued so many biomarker investigations. A previous study from our laboratory (12) is, to the best of our knowledge, the first report describing the concept of SELDI protein profiling as a potential diagnostic approach. This study observed that the selection of a combination of multiple proteins resolved by SELDI dramatically improved the detection rate of early-stage bladder cancer compared with a single marker (*i.e.*, urine cytology). Although the differential analysis in this latter study was conducted by cluster analysis and laborious manual visual inspection of all spectra, it did, however, demonstrate the power of SELDI profiling to facilitate the discovery of better cancer biomarkers. Furthermore, it clearly illustrated the need for a bioinformatics algorithm to effectively deal with the high dimensionality of the SELDI data. Based on the results of this previous study, we have explored several different bioinformatics models to mine and analyze the large amounts of data generated from these clinical proteomic studies. The models have included purely biostatistical algorithms, genetic cluster algorithms, support vector machines, and decision classification trees. All have obtained between 83–90% accuracy in separating PCA from the noncancer (BPH/HM) samples.[8] The classification tree model was selected because it is easy to interpret and the results can be clearly presented compared with "black box" classifiers such as neural networks and biostatistical algorithms, specifically with regard to the problems associated with the deconvolution steps required in identifying the protein peaks used in the classifiers. With the decision tree algorithm, the protein peaks used in the classifier are easily attainable by examination of the rules, and these rules are easily validated by examination of the SELDI processed spectra. Further proof of concept that coupling an artificial intelligent learning algorithm to analyze SELDI profiling data has potential as a diagnostic test is the recent report describing the use of a modified genetic algorithm that achieved a PPV of 94% in differentiating ovarian cancer from benign ovarian disease and healthy unaffected women (18). The discriminator pattern for classification of ovarian cancer in the study of Petricoin *et al.* (18) consisted of five protein masses of 534, 989, 2111, 2251, and 2465 Da. Although they used hydrophobic chip chemistry, which might be expected to bind some different proteins than those that would bind to the IMAC-3Cu chip used in the present study, it is interesting to note that the masses are distinctly different from those used in the prostate classification system. This suggests that the SELDI protein fingerprint profiling approach is detecting different protein patterns for each type of cancer. Studies in progress in our laboratory strongly suggest that this may be the case. We have observed that SELDI profiles of breast cancer, ovarian cancer, bladder cancer, and leukemia are different from each other and from the prostate classification profile described in this report.[9] To assure the robustness of our diagnostic system, the prostate classification algorithm is being challenged with non-PCAs and non-prostate diseases to determine that the protein profiling classification algorithm is specific for PCA. A similar scheme will be required of any disease-specific classification system.

One of the goals of this study was to identify markers in the prostate proteome that could potentially be used for early detection of cancer. Ongoing studies in our laboratory evaluating longitudinal serum samples over a 5–10-year period suggested that PCA may be suspected 5 or more years earlier than by PSA testing.[10] If validated with a larger number of patients, such studies will support the SELDI classification system as an early diagnostic test. However, to effectively apply this classification system for early detection, it will be essential to identify other biomarkers that can distinguish the aggressive cancers, *i.e.*, clinically important cancers, from nonaggressive cancers. Current evidence suggests that preoperative serum PSA <10 ng/ml is not a useful biomarker for predicting the presence, volume, grade, or rate of postoperative failure (4, 19). Thus, there is an urgent need for a better biological marker than PSA and all its molecular forms have been able to provide. A marker proportional to the volume of Gleason grade 4/5 (undifferentiated cancer) represents a critical need to more logically direct therapy tailored to tumor biology. Studies are in progress in our

[8] G. L. Wright, Jr., O. J. Semmes, P. Barlett, and C. Harris, unpublished observations.
[9] G. L. Wright, Jr., A. Vlahou, C. Laronga, J. Marks, and O. J. Semmes, unpublished observations.
[10] G. L. Wright, Jr., P. F. Schellhammer, and B-L. Adam, unpublished observations.

laboratory to evaluate SELDI serum spectra of pre- and postprostatectomy samples from patients who, after treatment, have biochemical evidence for recurrent disease in an effort to identify the biomarkers or risk factors that signal an aggressive cancer.

The successful use of the prostate classification system described herein relies entirely on the protein fingerprint pattern of the nine masses. Because these masses were found to be reproducibly reliably detected, only the mass values are required to make a correct classification or diagnosis. Knowing their identities for the purpose of differential diagnosis is not required. However, because knowing their exact identities will be essential for understanding what biological role these peptide/proteins may have in the oncogenesis of PCA, potentially leading to novel therapeutic targets, efforts are under way to purify, identify, and characterize these protein/peptide biomarkers. Furthermore, knowing their identities will be essential for producing antibodies for development of either classical or SELDI immunoassays, similar to the single and multiplex formats we described previously for the quantitation of PSA and prostate-specific membrane antigen (12, 20). The SELDI immunoassay format provides an alternate platform for quantitation of multiple biomarkers.

The high sensitivity, specificity, PPV, and NPF obtained by the serum protein profiling approach presented in this study demonstrate that SELDI protein chip mass spectrometry combined with an artificial intelligence classification algorithm can both facilitate discovery[11] of better biomarkers for prostate disease and provide an innovative clinical diagnostic platform that has the potential to improve the early detection and differential diagnosis of PCA.

Acknowledgments

We thank Dr. Richard Drake for critical reading of the manuscript.

References

1. Howe, H. L., Wingo, P. A., Thun, M. J., Ries, L. A., Rosenberg, H. M., Feigal, E. G., and Edwards, B. K. Annual report to the nation on the status of cancer (1973 through 1998), featuring cancers with recent increasing trends. J. Natl. Cancer Inst. (Bethesda), *93:* 824–842, 2001.
2. Djavan, B., Zlotta, A., Kratzik, C., Remzi, M., Seitz, C., Schulman, C. C., and Marberger, M. PSA, PSA density, PSA density of transition zone, free/total PSA ratio, and PSA velocity for early detection of prostate cancer in men with serum PSA 2.5 to 4.0 ng/mL. Urology, *54:* 517–522, 1999.
3. Pannek, J., and Partin, A. W. The role of PSA and percent free PSA for staging and prognosis prediction in clinically localized prostate cancer. Semin. Urol. Oncol., *16:* 100–105, 1998.
4. Stamey, T. A., Johnstone, I. M., McNeal, J. E., Lu, A. Y., and Yemoto, C. M. Preoperative serum prostate specific antigen levels between 2 and 22 ng/ml correlate poorly with post-radical prostatectomy cancer morphology: prostate specific antigen cure rates appear constant between 2 and 9 ng/ml. J. Urol., *167:* 103–111, 2002.
5. Srinivas, P. R., Srivastava, S., Hanash, S., and Wright, G. L., Jr. Proteomics in early detection of cancer. Clin. Chem., *47:* 1901–1911, 2001.
6. Adam, B-L., Vlahou, A., Semmes, O. J., and Wright, G. L., Jr. Proteomic approaches to biomarker discovery in prostate and bladder cancers. Proteomics, *1:* 1264–1270, 2001.
7. Chong, B. E., Hamler, R. L., Lubman, D. M., Ethier, S. P., Rosenspire, A. J., and Miller, F. R. Differential screening and mass mapping of proteins from premalignant and cancer cell lines using nonporous reversed-phase HPLC coupled with mass spectrometric analysis. Anal. Chem., *73:* 1219–1227, 2001.
8. Ferrari, L., Seraglia, R., Rossi, C. R., Bertazzo, A., Lise, M., Allegri, G., and Traldi, P. Protein profiles in sera of patients with malignant cutaneous melanoma. Rapid Commun. Mass Spectrom., *14:* 1149–1154, 2000.
9. Keough, T., Lacey, M. P., Fieno, A. M., Grant, R. A., Sun, Y., Bauer, M. D., and Begley, K. B. Tandem mass spectrometry methods for definitive protein identification in proteomics research. Electrophoresis, *21:* 2252–2265, 2000.
10. Merchant, M., and Weinberger, S. R. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis, *21:* 1164–1177, 2000.
11. Kuwata, H., Yip, T. T., Yip, C. L., Tomita, M., and Hutchens, T. W. Bactericidal domain of lactoferrin: detection, quantitation, and characterization of lactoferricin in serum by SELDI affinity mass spectrometry. Biochem. Biophys. Res. Commun., *245:* 764–773, 1998.
12. Wright, G. L., Jr., Cazares, L. H., Leung, S-M., Nasim, S., Adam, B. L., Yip, T. T., Schellhammer, P. F., and Vlahou, A. Proteinchip surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel proteomic technology for detection of prostate cancer biomarkers in complex protein mixtures. Prostate Cancer Prostatic Diseases, *2:* 264–267, 1999.
13. Vlahou, A., Schellhammer, P. F., Mendrinos, S., Patel, K., Kondylis, F. I., Gong, L., Nasim, S., and Wright, J. G., Jr. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. Am. J. Pathol., *158:* 1491–1502, 2001.
14. Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. Classification and Regression Trees. Belmont, CA: Wadsworth International Group, 1984.
15. Pepe, M. S. Receiver operating characteristic methodology. J.A.S.A., *95:* 308–311, 2000.
16. Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. Bayesian Data Analysis. London: Chapman & Hall, 1995.
17. Coley, C. M., Barry, M. J., Fleming, C., and Mulley, A. G. Early detection of prostate cancer. Part I: prior probability and effectiveness of tests. Ann. Intern. Med., *126:* 394–406, 1997.
18. Petricoin, E. F., III, Ardkani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M., Mills, G. B., Simone, C., Fishman, D. A., Kohn, E. C., and Liotta, L. A. Use of proteomic patterns in serum to identify ovarian cancer. Lancet, *359:* 572–577, 2002.
19. Stamey, T. A. Preoperative serum prostate specific antigen (PSA) below 10 μg/L predicts neither the presence of prostate cancer nor the rate of postoperative PSA failure. Clin. Chem., *47:* 631–634, 2001.
20. Xiao, Z., Adam, B-L., Cazares, L. H., Clements, M. A., Davis, J. W., Schellhammer, P. F., Dalmasso, E. A., and Wright, Jr., G. L. 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.

[11] L. H. Cazares, B-L. Adam, M. D. Ward, S. Nasim, P. F. Schellhammer, O. J. Semmes, and G. L. Wright, Jr. Normal, benign, preneoplastic, and malignant prostate cells have distinct protein expression profiles resolved by SELDI mass spectrometry, submitted for publication.