

Diagnosis of Pancreatic Cancer Using Serum Proteomic Profiling¹

Sudeepa Bhattacharyya*, Eric R. Siegel[†], Gloria M. Petersen[‡], Suresh T. Chari[§], Larry J. Suva* and Randy S. Haun[¶]

*Center for Orthopaedic Research, Department of Orthopaedic Surgery, University of Arkansas for Medical Sciences, Little Rock, AR, USA; [†]Department of Biometry, University of Arkansas for Medical Sciences, Little Rock, AR, USA; Divisions of [‡]Epidemiology and [§]Gastroenterology and Hepatology, Mayo Clinic Rochester, Rochester, MN, USA; [¶]Department of Pathology, University of Arkansas for Medical Sciences, Little Rock, AR, USA

Abstract

In the United States, mortality rates from pancreatic cancer (PCa) have not changed significantly over the past 50 years. This is due, in part, to the lack of early detection methods for this particularly aggressive form of cancer. The objective of this study was to use high-throughput protein profiling technology to identify biomarkers in the serum proteome for the early detection of resectable PCa. Using surface-enhanced laser desorption/ionization mass spectrometry, protein profiles were generated from sera of 49 PCa patients and 54 unaffected individuals after fractionation on an anion exchange resin. The samples were randomly divided into a training set (69 samples) and test set (34 samples), and two multivariate analysis procedures, classification and regression tree and logistic regression, were used to develop classification models from these spectral data that could distinguish PCa from control serum samples. In the test set, both models correctly classified all of the PCa patient serum samples (100% sensitivity). Using the decision tree algorithm, a specificity of 93.5% was obtained, whereas the logistic regression model produced a specificity of 100%. These results suggest that high-throughput proteomics profiling has the capacity to provide new biomarkers for the early detection and diagnosis of PCa.

Neoplasia (2004) 6, 674–688

Keywords: SELDI, surface-enhanced laser desorption/ionization, mass spectrometry, proteomics, early detection.

radiation exposure) risk factors [3–7]. Thus, it has been difficult to define an “at-risk” population for PCa screening.

As a further complication, detection of adenocarcinoma at an early, treatable stage is difficult because the disease lacks specific symptoms. Early symptoms of pancreatic carcinoma including weight loss, anorexia, epigastric discomfort, and back pain are often nonspecific and vague, so diagnosis may be considerably delayed [8,9]. Furthermore, primary tumors of the pancreas are relatively inaccessible to detection by routine physical examination. Modern technological advances including ultrasonography, dynamic computed tomography (CT scan), magnetic resonance imaging (MRI), angiography, endoscopic retrograde cholangiopancreatography (ERCP), and endoscopic ultrasonography have certainly aided the diagnosis and staging of pancreatic carcinoma [7,10]. These new technologies, however, have not materially influenced survival from PCa [10]. Although a wide variety of tumor-associated antigens have been evaluated as markers for screening and diagnosis of PCa, most have proven ineffective due to their low sensitivity and cross-reactivity with other tumors. Thus, the poor prognosis of patients is primarily due to the fact that most patients present, and are treated, in the terminal stage of the disease.

Despite the efforts of many groups, tumor markers that are specifically expressed in PCa have yet to be identified. Currently, the tumor markers carcinoembryonic antigen (CEA) and CA19-9, and mutations in the oncogene *Ki-ras* have been employed in the staging and diagnosis of pancreatic neoplasms [11]. It has been found, however, that CA19-9 is effective in discriminating between PCa and chronic pancreatitis, but not between PCa and other digestive cancers [12]. Although the diagnostic accuracy of CA19-9 appears superior

Introduction

Since 1950, the annual incidence of pancreatic cancer (PCa) in the United States has increased from 5.3 to 9.2 cases per 100,000 population. The number of new cases of pancreatic adenocarcinoma diagnosed each year (~27,000 cases) essentially equals the mortality rate from this disease (~26,000 deaths) [1,2]. Although a number of studies indicate that the risk for PCa of cigarette smokers is more than twice that for nonsmokers, in general, there is little evidence for other extrinsic (e.g., alcohol or coffee consumption, occupational exposure to carcinogens, and

Abbreviations: AUC, area under the curve; CART, classification and regression tree; ICC, intraclass correlation coefficient; MS, mass spectrometry; PCa, pancreatic cancer; ROC, receiver-operating characteristic; SELDI, surface-enhanced laser desorption/ionization; TIC, total ion current; TOF, time of flight

Address all correspondence to: Randy S. Haun, PhD, Department of Pathology, University of Arkansas for Medical Sciences, 4301 West Markham Street, No. 753, Little Rock, AR 72205-7199. E-mail: HaunRandyS@uams.edu

¹This work was supported, in part, by a grant from the Arkansas Tobacco Settlement (R.S.H.). Received 30 March 2003; Revised 8 June 2004; Accepted 18 June 2004.

to that of other tumor markers currently available, these findings are valid only for patients with advanced cancers [11–17]. CEA, a membrane glycoprotein originally described as a tumor-associated colon cancer antigen, is now widely used in clinical practice as a tumor marker [18]. CEA and other tumor markers appear to have far less sensitivity and specificity than CA19-9 in PCa. Sensitivity and specificity for CEA have been reported to range from 59% to 71% and from 63.9% to 66.4%, respectively [14,16]. Using an upper limit of normal of 37 U/ml, assays for CA19-9 have been reported to range from 79.4% to 89.2% sensitivity and from 72.5% to 90% specificity [14,16,17]. Using a higher cutoff of 1000 U/ml, the specificity of CA19-9 approaches 100%; however, the sensitivity decreases to 24.3% to 41% [16,17]. CEA is normally present in the human embryo and is found only in minute amounts in healthy adults; however, it can be highly expressed by malignant cells throughout the gastrointestinal tract, as well as tumor cells arising from diverse locations such as the breast, ovary, and lung [11]. Thus, the clinical utility of CA19-9 is viewed by some clinicians as most valuable in patients presenting with signs and symptoms of a chronic pancreatic disorder, rather than as a screening test to detect PCa in asymptomatic individuals [15,17].

Detection of cancer-derived gene products from biologic fluids is an important emerging approach to the diagnosis of malignant diseases. Due to the lack of any specific or sensitive diagnostic test for the early stages of PCa, there is a critical need for tumor markers to aid in the early detection of this disease. Recent advances in techniques used to generate “fingerprints” of cancer cells and identify proteins elicited by tumors based on mass spectrometry have yielded new biomarkers for the early detection of cancers [19–22]. In particular, surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) [23–25] has been successfully applied to the identification of serum biomarkers for the detection of breast [26], liver [27], ovarian [28,29], and prostate cancers [30,31]. The goal of this study, therefore, was to use high-throughput SELDI-TOF MS to identify directly the signature of serum proteins in patients with pancreatic tumors compared with patients without malignancies. We expect that new biomarkers, which can be developed for the detection of this deadly disease at a treatable (resectable) stage, will be discovered.

Materials and Methods

Sample Collection and Preparation

Whole blood (5 ml) was collected in a serum separator tube (SST) from patients at the University of Arkansas for Medical Sciences Hospital and Clinics (Little Rock, AR) using a protocol approved by the institutional IRB. Specimens were collected from PCa patients and individuals with no evidence of PCa. A self-administered questionnaire was collected from each patient pertaining to gender, age, smoking and alcohol usage, and medical history. For the separation of serum, the blood in the SST tube was allowed to clot for 15

minutes and then centrifuged for 10 minutes at 1500g. The serum from the SST was then aliquoted into cryovials and immediately frozen at -70°C . Serum samples from PCa patients were obtained from the Biospecimen Core in the Pancreas Clinic at the Mayo Clinic Rochester (Rochester, MN). Samples processed by the Mayo Clinic Molecular Genetics Laboratory were collected in an EDTA lavender or ACD yellow top tube, centrifuged at 3000 rpm for 10 minutes, aliquoted into tubes, and frozen at -70°C . All serum samples were labeled with a unique identifier to protect the confidentiality of the patient. None of the samples was thawed more than twice before analysis.

The mean age (\pm SD) of the PCa patients was 66.4 ± 12.2 vs 53.5 ± 15.7 years among the control patients ($P < .0001$). Thirty (61%) of the 49 cancer patients were male, compared to 15 (28%) of 54 control patients ($P = .0006$). The 103 serum samples were randomly divided into two groups in a 2:1 ratio to yield a training set comprised of 36 patients with no evidence of PCa, designated as “normal,” and 33 patients diagnosed with PCa and a test set comprised of 34 serum samples from both normal ($n = 18$) and PCa ($n = 16$) patients. Staging information was available on 45 of 49 PCa patients with a distribution of one stage I, 12 stage II, 13 stage III, and 19 stage IV patients. Included in the self-administered medical histories of the non-PCa samples were seven patients with a history of other cancers (five multiple myelomas, one chronic lymphoblastic leukemia, and one testicular cancer). A quality control (QC) sample was also prepared by pooling serum from eight healthy male and eight healthy female age-matched individuals.

Serum Fractionation

To increase the detection of a larger number of peaks as well as to alleviate signal suppression effects on lower-abundance proteins from highly abundant proteins such as albumin, serum samples were fractionated into six fractions containing proteins separated roughly on the basis of their isoelectric points. Serum samples were loaded onto each well of a 96-well filter plate prefilled with an anion exchange sorbent (Serum Fractionation kit; CIPHERGEN Biosystems, Fremont, CA) and eluted in a stepwise pH gradient using a BIOMEK 2000 liquid-handling robot as described by the manufacturer. The fraction containing the flow-through plus proteins eluted with pH 9 buffer, which yields the most protein peaks on the IMAC surface (S.B., unpublished observation) [34], was selected for analysis. Each serum sample was diluted 10-fold during fractionation in 50 mM Tris-HCl, pH 9, containing 0.1% nonionic detergent.

ProteinChip Array Analysis

For analysis of fractionated sera, samples were further diluted 1:5 in phosphate-buffered saline (PBS) and applied in duplicate onto each well of a 192-well bioprocessor containing 16-spot IMAC-30 chips (CIPHERGEN Biosystems) previously activated with 100 mM CuSO_4 . The bioprocessor was sealed and incubated with the samples for an hour, with vigorous agitation on a Micromix 5 platform shaker. A pooled QC sample prepared in the same manner was applied to

duplicate spots on each chip used in each experiment as a reproducibility control. The excess sera mixtures were discarded and the chips were washed three times with PBS. The chips were then washed with deionized water, removed from the bioprocessor, and air-dried for 20 minutes. A saturated solution of sinapinic acid (0.5 μ l) in 50% acetonitrile, 0.5% trifluoroacetic acid was applied to each spot on the chip surfaces. The array surface was allowed to dry for 10 minutes before another application of 0.5 μ l of the sinapinic acid solution.

Data Acquisition

ProteinChip arrays were read by a PBS-II C mass analyzer (Ciphergen Biosystems) and the spectral data were acquired using Ciphergen Biosystems' ProteinChip software version 3.1. The TOF spectra were generated by averaging 156 laser shots in the positive mode with a laser intensity of 180, detector sensitivity of 8, and a focus lag time of 782 nanoseconds. The data acquisition parameters were optimized to detect peaks in the range of 2 to 20 kDa, as this range contained the majority of the resolved protein/peptide peaks. Mass accuracy was calibrated using the all-in-one peptide and all-in-one protein molecular weight standards (Ciphergen Biosystems).

Data Analysis

The serum samples were analyzed in two batches. Peaks were baseline-corrected and the spectra were normalized by total ion current (TIC), the sum of ion intensities between the 2 and 20 kDa mass range applying the normalization coefficient of the initial set of samples analyzed to normalize the second set. Spectra for which the normalization factors were either > 2 or < 0.5 were discarded. Using Ciphergen Biosystems' Biomarker Wizard tool available with the Protein Chip software, peaks consistently present across a minimum of 20% of the spectra with a signal-to-noise ratio ≥ 2.5 and present within a mass window of 0.3% were detected and clustered in the set. Corresponding peaks were likewise identified in the spectra from the second set of samples using the clustering data from the initial set.

Univariate Analysis

For each peak, the median patient-averaged intensity was calculated for the normal and cancer groups. The difference in group medians was reported as a ratio, the fold change, and assessed for statistical significance through the Wilcoxon rank-sum test with t approximation. Multiple comparison adjustment of P values was through the Stepdown Permutation procedure of Westfall and Young [32] using 100,000 random permutations of class labels. A peak was deemed to show a statistically significant difference in group medians if its multiple comparison-adjusted P value was less than .05.

Reproducibility

To examine the uniformity of spectral data collected within and between experiments, a sample of the QC serum was spotted in duplicate on each chip. The resulting spectra were

normalized, and the intensities of each peak identified by the Ciphergen Biosystems software were measured. To measure variation between spots within a chip, the Spearman correlation was calculated for the two QC samples on each chip and the median correlation was determined for the 10 chips used in the training set. For the interchip correlation, the average intensities for the two QC samples per chip were determined, the Spearman correlation was calculated for each of five randomly selected pairs of chips used in the training set, and the median correlation of the pairs was reported.

Intracluster Correlation Analysis

After TIC normalization, a total of 164 useable spectra was obtained from 103 patients, with 61 patients contributing pairs of spectra and 42 patients contributing single spectra. It was expected that SELDI peaks from spectra paired by patient would show appreciable correlation of intensities. To determine the amount of correlation, the paired spectra were subjected to variance components analysis on a peak-by-peak basis using a random effects model with "patient" as the random effect. In this manner, the within-patient and between-patient variance terms were obtained for each peak. The intracluster correlation coefficient (ICCC) was calculated as the ratio of the between-patient variance to the sum of the variance terms. For the 37 peaks with an m/z ratio between 2000 and 20,000, the ICCCs had a median value (interquartile range) of 85% (79–89%). Because of the high ICCCs, the paired spectra from each patient were averaged together on a peak-by-peak basis for subsequent logistic regression analyses.

Multivariate Logistic Regression Classification

In multivariate model development, a weighted logistic regression was used, in which the 42 patients contributing single spectra were given a weight of 1.00, whereas the 61 patients contributing the average of two spectra were given a weight equal to 2 minus the median ICCC (about 1.15). Weights were then multiplied by a constant so that their sum would be equal to the sample size. Our approach to developing a classification model was guided by the recognition that the results of automated variable selection procedures can depend heavily on how the patients are allocated into training and test sets. We sought for our final model a set of peaks that would be insensitive to such allocation effects. To accomplish this, we developed our classification model in several stages.

In stage 1, the 103 patients (49 cancer; 54 normal) were randomly divided into a training set of 52 patients (25 cancer) and a test set of 51 patients (24 cancer). This process was repeated 10,000 times to generate 10,000 random divisions of the patients into training and test sets.

In stage 2, logistic regression was applied to the training set of each random division of stage 1, with SELDI peaks chosen by forward selection under an entry threshold of 0.1, and the resulting model was then applied to the test set of that random division. The response variable was the cancer *versus* normal classification, and the goal was to predict

correctly the class of test set members. Cross-validation and test set class prediction probabilities were recorded along with model parameter estimates. In this manner, we collected 10,000 logistic regression models generated by forward selection on randomly chosen training sets, and 10,000 associated classification probabilities for each patient, approximately half of which were test set class prediction probabilities. The frequency of each peak's incorporation into a forward selection model was calculated, and the five most frequently incorporated peaks ("the five good peaks") were singled out for further modeling in stage 3. Additionally, probabilities of prediction into the correct class were summed across all test set patients within a random division. Final models whose sums of correct class prediction probabilities were near the maximum of 51 showed three frequent patterns of peak selection; the most frequent of these ("the four-peak pattern") was also singled out for further modeling in stage 3.

In stage 3, logistic regression using the five good peaks or the four-peak pattern was applied without variable selection to the training set of each random division of stage 1, and the resulting model was then applied to the test set of that random division. As in stage 2, the goal was to predict correctly the class of test set members. Cross-validation and test set class prediction probabilities were recorded along with parameter estimates for models using the five good peaks, and the same was done for models using the four-peak pattern. In this manner, we generated a distribution of estimated parameter vectors for both model types, and a distribution of classification probabilities for each patient. For each model type, the probabilities of correct class prediction were summed across all test set patients within a random division for use in stage 4.

In stage 4, the sums of correct class prediction probabilities were assigned fractional ranks under the "ties=high" rule; this was done separately for the four-peak-pattern models, the five-good-peak models, and the forward selection models. Performance of the three types of models across 10,000 random divisions was then compared visually by plotting sums of correct class prediction probabilities against their fractional ranks. To provide a second measure of performance for the three types of models, each patient's test set probabilities for cancer prediction were averaged together to give an average Prob(Cancer) based on the number of times the patient appeared in 10,000 test sets. Receiver-operating characteristic (ROC) curves were then calculated for each type of model. The area under the curve (AUC) was calculated and used to compare the three types of models for average performance on 10,000 random divisions of the data into training and test sets.

On the basis of stage 4 results, it was determined that models based on the four-peak pattern, consisting of SELDI peaks having median m/z ratios of 3966.8, 3983.1, 8951.7, and 7787.2, were far more likely to classify well, regardless of random allocation into training and test sets, when compared to models based on either the five good peaks or forward selection. To derive a final multivariate classification model, the 103 patients (49 cancer) were divided randomly one more time into a training set of 69 patients (33 cancer)

and a test set of 34 patients (16 cancer). The four-peak pattern was trained on the training set and applied to the test set, and sensitivities and specificities of class prediction were calculated under resubstitution, cross-validation, and test set classification. The likelihood ratio test was used to determine whether it was necessary to covariate-adjust the final classification model for the age and gender imbalance between cancer and normal patients.

Classification and Regression Tree (CART) Model

The Biomarker Pattern software (version 4.0.1; Ciphergen Biosystems) implements the CART statistical procedure described by Breiman et al. [33] to build a decision tree that can classify a sample set into a given number of groups. Classification trees were developed using a variety of predictors (peaks) and program parameters to build trees. Tree building was then repeated to yield the best prediction success with the lowest error cost. The optimal tree was produced using a Gini power of zero, a minimum parent node size of four, a 10-fold cross validation, and univariately significant peaks with a median intensity > 10 in either the cancer or normal serum samples as the input predictors.

Results

Serum Protein Profiles

To begin to screen for potential serum biomarkers and identify a unique signature of serum proteins for early detection of resectable PCa using the SELDI-TOF MS technique, sera from 49 PCa patients and 54 control patients were fractionated by anion exchange chromatography. In this study, proteins that eluted with pH 9 buffer (flow-through + fraction 1) were applied in duplicate to IMAC3 ProteinChips (Ciphergen Biosystems) and bound proteins were detected with a ProteinChip Reader after laser desorption/ionization. After normalization, 164 spectra were compiled and mass peaks with mass-to-charge ratios (m/z) between 2000 and 20,000 and a signal-to-noise ratio > 2.5 were identified, clustered, and analyzed. As shown in Figure 1, differences in protein profiles between serum samples collected from PCa patients *versus* controls were readily detected using SELDI-TOF MS.

ProteinChip Reproducibility

To examine the variation in data collection within and between the ProteinChips used in these experiments, a QC sample representing a pool of sera from age-matched male and female control subjects was applied in duplicate to each ProteinChip. For each of the 10 chips used in the initial analysis, the Spearman correlation was determined for the QC samples on each chip, resulting in a median intrachip correlation of 0.89. To examine the interchip correlation, the average intensities of the two QC spectra on each chip were determined, and the Spearman correlation was calculated for five randomly selected pairs of the 10 chips used in the initial set. The median interchip correlation for the five pairs was 0.95.

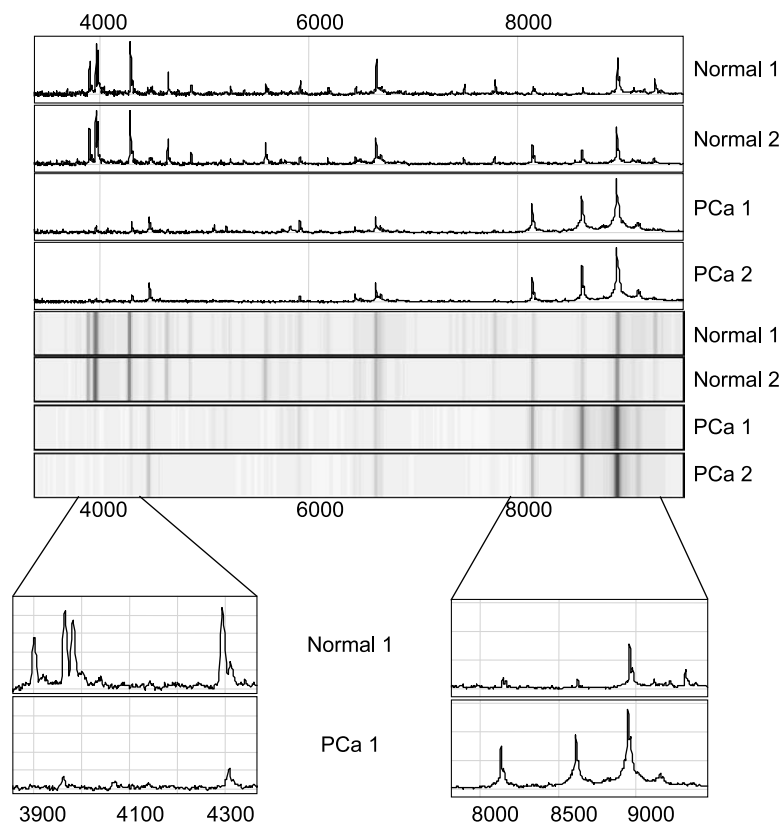


Figure 1. Difference in SELDI spectra between normal and cancer serum samples. Upper panel: A portion of spectra from normal and PCa serum samples is depicted as tracings in GelView (darker shades indicate higher mass peak intensities). Lower panel: An expanded portion of tracings from a PCa patient serum and an unaffected individual is displayed, highlighting the differences in mass ions detected in the two patient populations.

Univariate Analysis

Of the 37 SELDI peaks with median m/z ratio between 2000 and 20,000, 20 showed a statistically significant difference ($P < .05$) in median intensities between the cancer and control samples (Table 1). Table 1 shows the fold change of increase or decrease, along with raw and adjusted P values. An adjusted P value of zero means that the corresponding raw P value was lower than any produced by chance from 100,000 random permutations of class labels. Eight of these specific protein peaks were elevated in PCa serum samples (1.66- to 2.24-fold, PCa versus normal) and 12 peaks were higher in sera from unaffected individuals (1.96- to 10.86-fold, control versus PCa).

Multivariate Logistic Regression Classification

Figure 2 shows the proportion of times each peak was incorporated under forward selection into classification models trained on 10,000 randomly chosen training sets of 52 patients (25 cancer). Only the 25 most frequently selected peaks are shown, although every peak was selected at least once. Median m/z ratios (incorporation rates \pm standard errors of rate) for the five peaks with highest multivariate usage are: 3966.8 (92.8% \pm 0.26%), 3983.1 (65.6% \pm 0.48%), 4309.4 (32.8% \pm 0.47%), 8951.7 (21.0% \pm 0.41%), and 5592.5 (17.1% \pm 0.38%). These five peaks became “the five good peaks” of additional classification modeling. The 6th to 10th peaks (median m/z of 9157.9, 3902.2, 4479.5,

8620.7, and 7787.2) had usage rates ranging from 12.3% to 10.4%, the 11th to 25th peaks had usage rates ranging from 4.62% to 1.00%, and the 12 peaks not shown in the figure had usage rates ranging from 0.90% to 0.12%. The two predominant features of Figure 2 are: 1) the sharp but smooth drop in incorporation rates from 90% to 10% for the first 10 peaks, and 2) the discontinuity in the trend between the 10th and 11th peaks. When the 10,000 forward selection models were applied to their test sets, 128 yielded sums of correct class prediction probabilities within 0.1 of the maximum value of 51. The peaks at 3966.8 and 3983.1 appeared in all 128; additional peaks at 8951.7 and 7787.2 appeared in more than half. These four peaks were the 1st, 2nd, 4th, and 10th most frequently incorporated peaks under forward selection; they became “the four-peak pattern” of additional classification modeling.

In additional classification modeling, the same 10,000 randomly chosen training sets were used to train classification models containing either the five good peaks or the four-peak pattern; the resulting models of each type were then applied to their test sets to obtain class prediction probabilities. For each test set and model type, the probabilities of prediction into the correct class were summed across patients, and the sums given fractional ranks. For each patient, the probabilities of prediction into the cancer class were averaged across test sets in which the patient appeared, and the average probabilities were used to generate ROC curves. Figure 3 shows the results of plotting sums of

Table 1. Significant Peaks by Univariate Analysis.

m/z	Median Cancer	Median Normal	Cancer Effect	Fold Change	Raw* WRS P values	Adjusted† P values	Peaks Used	
							CART	LR
3,276.1	0.488	1.948	Down	3.99	1.31E-08	0		
3,902.2	2.583	11.608	Down	4.49	2.45E-20	0	X	
3,966.8	3.291	25.679	Down	7.80	3.58E-23	0	X	X
3,983.1	1.830	15.466	Down	8.45	5.93E-16	0	X	X
4,295.4	1.721	18.692	Down	10.86	3.72E-19	0		
4,309.4	3.341	6.560	Down	1.96	4.43E-05	0.00122		
4,479.5	4.603	2.492	Up	1.85	5.04E-05	0.00126		
4,651.3	1.105	7.629	Down	6.91	1.04E-12	0		
5,592.5	0.210	1.461	Down	6.95	6.94E-12	0		
6,453.2	2.146	5.868	Down	2.73	0.0026	0.04274		
6,648.1	6.124	18.273	Down	2.98	0.00032	0.00649		
7,487.2	0.867	2.485	Down	2.87	5.91E-11	0		
7,787.2	1.437	4.125	Down	2.87	2.55E-07	0.00003		X
8,620.7	16.195	7.768	Up	2.08	5.94E-08	0.00001	X	
8,951.7	29.137	15.360	Up	1.90	1.37E-08	0	X	X
8,966.2	15.177	7.188	Up	2.11	1.88E-09	0		
9,157.9	4.067	1.816	Up	2.24	2.83E-11	0		
11,498.4	0.354	0.213	Up	1.66	0.0026	0.04274		
11,654.9	0.377	0.178	Up	2.12	0.00069	0.01349		
11,711.9	0.393	0.215	Up	1.83	0.00071	0.01349		

*Wilcoxon rank-sum test with *t* approximation.

† P values adjusted for multiple comparisons using stepdown permutation procedure [32].

probabilities against their fractional ranks for the forward selection models, the five-good-peak models, and the four-peak-pattern models. The sum of probabilities per test set can be interpreted as the expected number of correctly classified patients in that test set; it therefore has a maxi-

imum value of 51. Under forward selection, 128 models (1.28%) yielded an expected number within 0.1 of the maximum value, whereas 9382 models (93.82%) yielded an expected number below 49.9. Using the five good peaks, 144 models (1.44%) yielded an expected number within 0.1

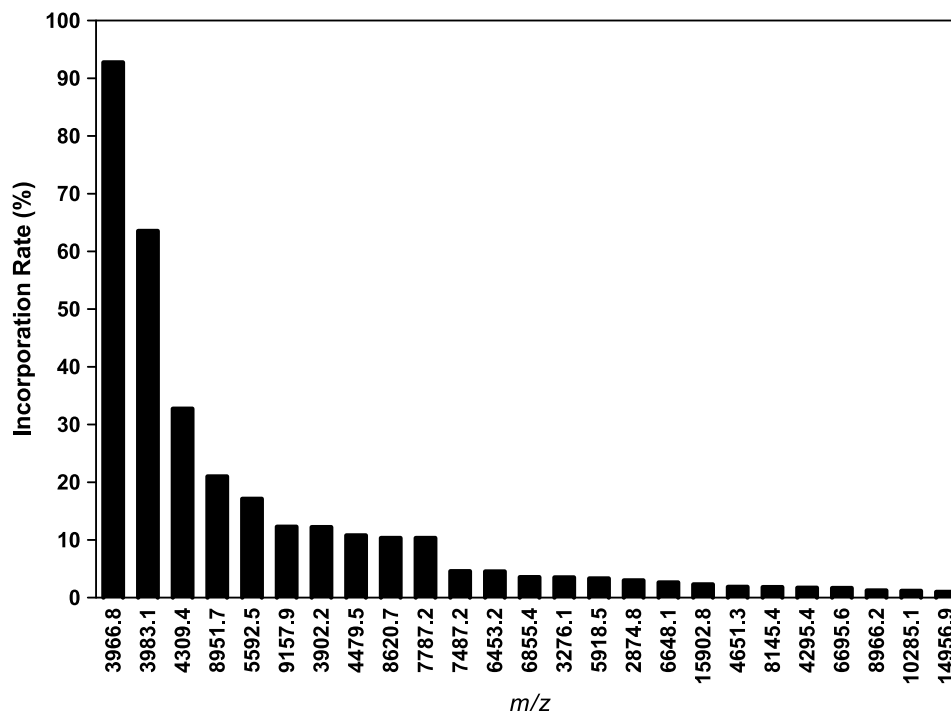


Figure 2. Rate of incorporation into multivariate logistic classification models. The rate of incorporation was based on 10,000 random divisions of the data from 52 patients into training and test sets of approximately equal size. Each training set was used to develop a classification model via multivariate logistic regression using forward selection with an entry threshold of $P = .10$. The incorporation rate for each peak is thus the proportion of times it was selected into 10,000 multivariate classification models trained on random samples of half of the data. Only the 25 most frequently selected peaks are shown.

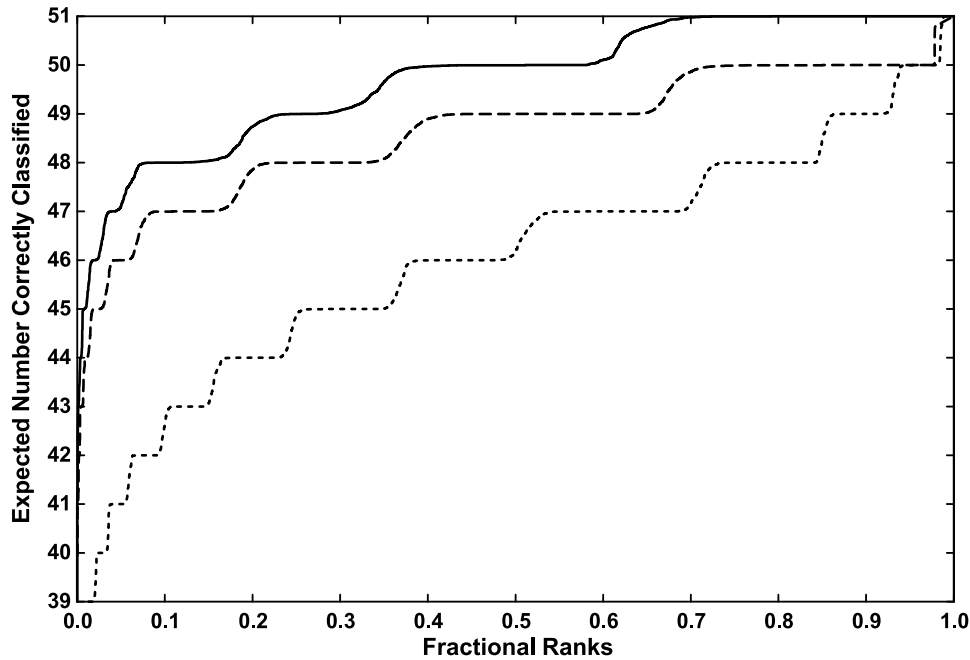


Figure 3. Plots of sums of probabilities against their fractional ranks for the forward selection models. The results of plotting sums of probabilities against their fractional ranks for the four-peak-pattern (solid line), five-peak-pattern (dashed line), and forward selection (dotted line) models. The sum of probabilities per test set can be interpreted as the expected number of correctly classified patients in that test set; it therefore has a maximum value of 51.

of the maximum value, whereas 7041 models (70.41%) yielded an expected number below 49.9. Using the four-peak pattern, 3280 models (32.80%) yielded an expected number within 0.1 of the maximum value, and only 3702 models (37.02%) yielded an expected number below 49.9. Figure 4 shows the ROC curves generated from the aver-

age of cancer prediction probabilities for each patient when they appeared in a test set. The AUCs are 0.9845 for the forward selection models, 0.9962 for the five-good-peak models, and 0.9996 for the four-peak-pattern models; an AUC of 1.0000 would denote a correct classification of every patient in every one of their appearances in a test

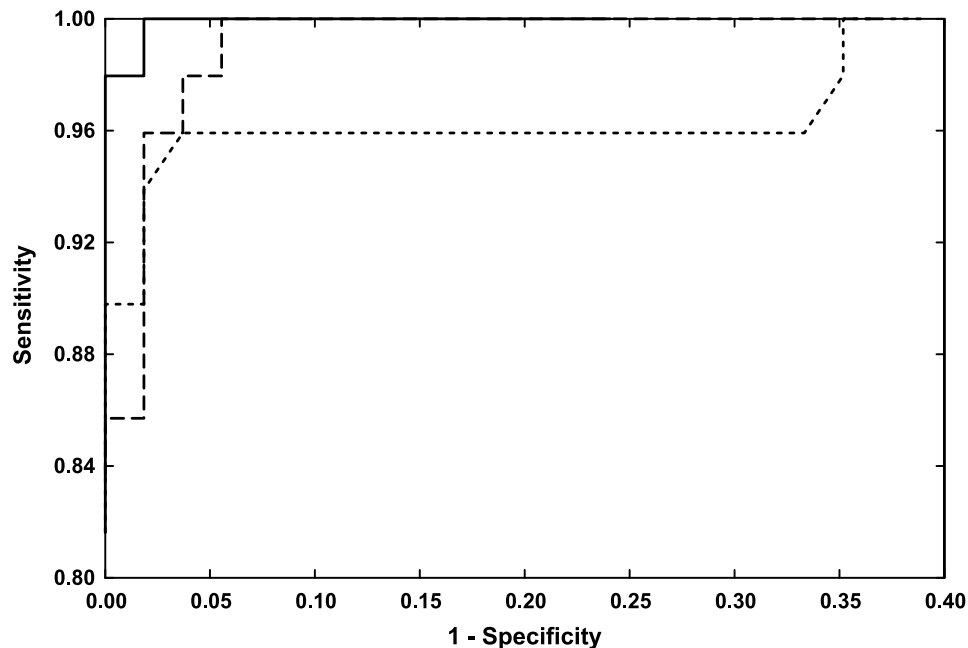


Figure 4. ROC curves generated from the average of cancer prediction probabilities for each patient. ROC curves generated from the average probability of cancer based on each patient's appearances in 10,000 random test sets. The AUC for the four-peak-pattern models (solid line) is 0.9996, for the five-peak-pattern models (dashed line) is 0.9962, and for the forward selection models (dotted line) is 0.9845.

set. These results indicate that models generated from the four-peak pattern tend far more often to classify test set patients correctly, compared to models generated from either the five good peaks or from forward selection on all 37 peaks.

To derive parameter estimates for the final logistic regression classification model, the 103 patients were randomly divided one more time in a 2:1 ratio, so that the training set had 69 patients (33 cancer) and the test set had 34 patients (16 cancer). The training set was then subjected to logistic regression using the four-peak pattern, and the resulting model was then tested on the test set. Table 2 shows the parameter estimates and classification rule for the four-peak-pattern model trained on this particular training set, and Table 3 shows the sensitivity and specificity for the classification rule under resubstitution, cross-validation, and test set prediction. In additional analysis (not shown), we studied whether the confounding of cancer with age and gender in this data set would require their addition as covariates to the final classification model. We resubjected the training set to logistic regression using the four-peak pattern augmented with age and gender, and then used the likelihood ratio test to compare the augmented model to the model of Table 2. The comparison yielded a $P > .999$, strongly suggesting that covariate adjustment for age and gender is not needed.

CART Analysis

The Biomarker Patterns software (CIPHERGEN Biosystems) was used to identify a subset of mass peaks that could discriminate the patient sera. Initially, all 20 of the significant peaks identified by univariate analysis were evaluated as potential predictors using 111 spectra derived from the 69 serum samples initially used in the same training set used for the final logistic regression classification model. Each split was then rank-ordered on the basis of the quality-of-split criterion using the Gini rule (a measure of how well the splitting rule separates the classes contained in the parent node) [33]. Once a best split was found (e.g., node 1 in Figure 5), the algorithm repeated the search process for each child node, continuing recursively until further splitting was impossible or stopped. After the maximal tree was

Table 3. Classification Results.

	CART		Logistic Regression	
	Correct/Total	Percent	Correct/Total	Percent
<i>Training set resubstitution results</i>				
Sensitivity	55/57	96.5	33/33	100
Specificity	54/54	100	36/36	100
<i>Training set cross-validation results</i>				
Sensitivity	53/57	93.0	31/33	93.9
Specificity	51/54	94.4	35/36	97.2
<i>Classification results for test set</i>				
Sensitivity	22/22	100	16/16	100
Specificity	29/31	93.5	18/18	100

grown, smaller trees were examined by pruning away branches of the maximal tree, using cross-validation to estimate the error rate of the subtrees. Following this same procedure, a variety of possible predictors (peaks) and program parameters were used to build an optimal classification tree that yielded the lowest error cost. The optimal tree (i.e., the tree with the lowest error cost) was generated using the eight peaks with significant differences between the cancer and normal groups that had a median intensity > 10 in either the cancer or normal group as the input predictors (Table 1). From these initial eight peaks, five peaks were identified, which discriminated the control and PCa sera (Figure 5) with a sensitivity of 96.5% and a specificity of 100% (Table 3). Three of these five peaks ($m/z = 3966.8, 3983.1, \text{ and } 8951.7$) had also been identified in the multivariate logistic regression model as the best discriminators of PCa *versus* non-PCa sera (Table 1). Using a 10-fold cross-validation of the training set, a sensitivity of 93% and a specificity of 94.4% were obtained (Table 3). The decision matrix embodying the classification tree (Figure 5) was then applied to a test set of 53 spectra, representing 34 serum specimens. All of the 22 PCa spectra and 29 of 31 normal spectra were correctly classified, yielding a sensitivity and a specificity of 100% and 93.5%, respectively.

Group Comparisons

The intensities of the peaks used in the CART and logistic regression models were plotted and compared for individual spectra representing the control patients and PCa patients for whom staging information was available (Figure 6). The scatter plots reveal that the range of peak intensities for any of the individual peaks used to classify normal *versus* cancer overlaps between the normal samples and the PCa samples regardless of tumor stage. This overlap may reflect the historical failure to find a single marker that changes in serum as pancreatic tumors (or most other tumors) progress, and supports the notion that the successful discovery of new diagnostic biomarkers will require a panel of markers whose changes in serum levels are more subtle.

Discussion

Currently, there are no methods for the early detection of PCa. As a consequence, patients frequently present with

Table 2. Logistic Classification Model.

Parameter Name	Estimate of Coefficient	Likelihood Ratio Test of H_{null} : All Coefficients Equal Zero	
Intercept	173.8	Likelihood ratio chi square analysis	95.5993
MZ03966.8	-23.0981	Degrees of freedom	4
MZ03983.1	18.3736		
MZ08951.7	6.9714	P value	< .0001
MZ07787.2	-46.2511		

Classification equation: Score = 173.8 + (-23.0981*MZ03966.8) + (18.3736*MZ03983.1) + (6.9714*MZ08951.7) + (-46.2511*MZ07787.2), where "Score" equals the natural logarithm of the odds of being cancer, and "MZ0xxx.x" represents the intensity of the peak with the indicated m/z ratio. Classification decision rule: If Score is positive, classify as cancer; otherwise, classify as normal.

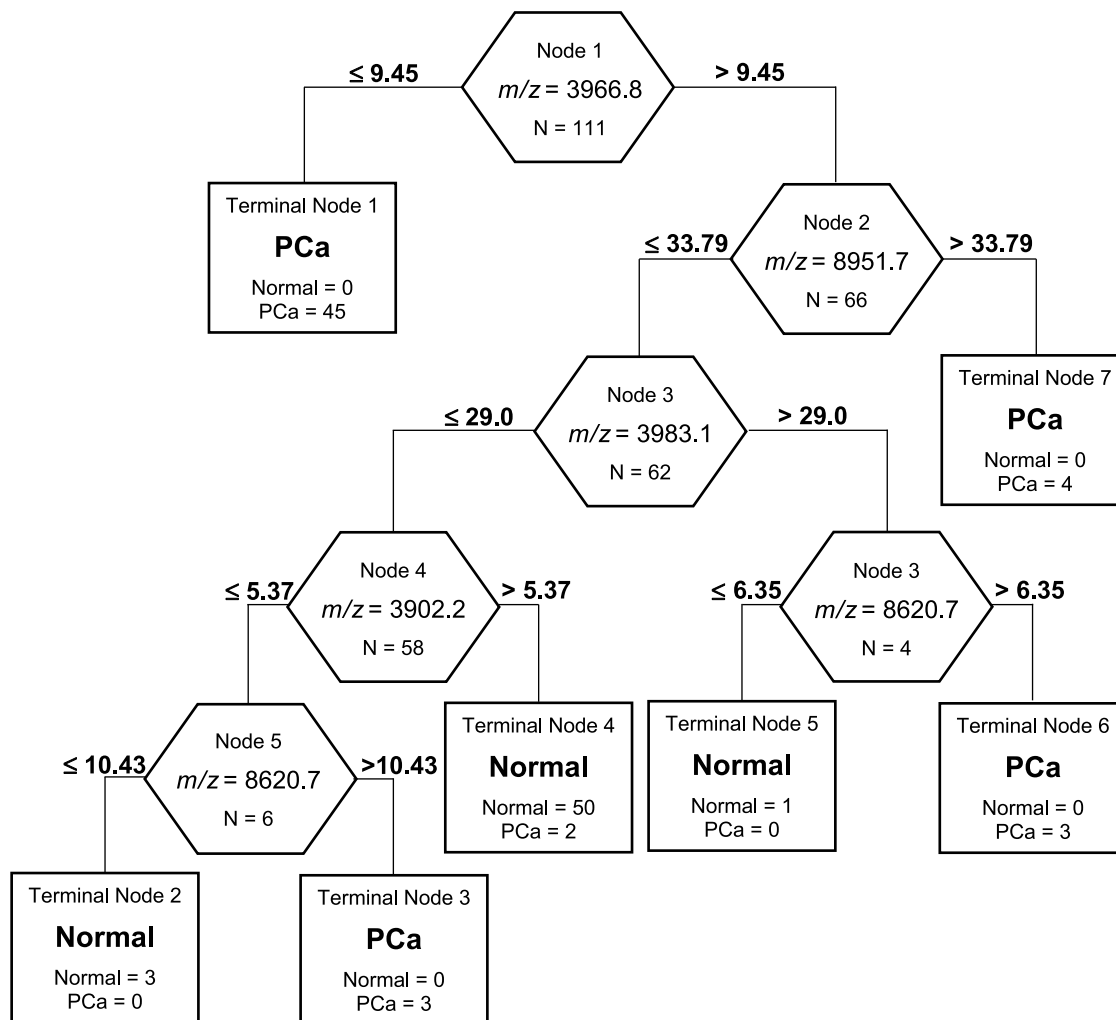


Figure 5. CART decision tree. Spectra from 69 serum samples were used as a learning set to generate a decision tree to distinguish between serum obtained from PCa patients and serum from unaffected individuals. Decision nodes (hexagons) represent individual mass peaks (m/z) and a threshold criterion for traversing the tree (top to bottom). Terminal nodes (squares) determine whether a sample is classified as normal or PCa. One PCa sample (represented by two spectra) was misclassified in the training set belonging to terminal node 4.

metastatic disease for which treatment is palliative rather than curative. The most widely used tumor-associated antigen for PCa, CA19-9, lacks adequate sensitivity and specificity and has limited utility as an indicator of early, localized pancreatic disease. As early surgery is the only curative intervention, there is a lingering need for better biomarkers for the early detection of resectable PCa. Recently, the use of SELDI-TOF MS for high-throughput profiling of serum proteins has gained attention as a facile method to identify panels of peptides and proteins that may be indicative of early disease for several cancers [26–31]. Using a similar approach, we have identified protein profiles in this study that were capable of distinguishing sera of PCa patients from unaffected individuals.

A total of 103 serum samples, representing 49 patients with PCa and 54 individuals with no evidence of PCa, were used to prepare protein profiles. To enhance the selection of mass peaks with biologic relevance, spectral data were analyzed by two independent multivariate methods (CART

and multivariate logistic regression). In each of these methods, a small subset of mass peaks that could readily distinguish the two sample populations was identified. A comparison of the features selected by these two models indicates that three of the peaks are used as primary discriminators in both models ($m/z = 3966.8$, 3983.1 , and 8951.7) (Table 1). This observation provides confidence that these mass peaks are disease-relevant classifiers.

The manner in which the aforementioned mass peaks found their way into our logistic regression final model gives us additional confidence that they are disease-relevant classifiers. The process by which these peaks were identified was similar to the Bootstrap selection procedure described by Koopmann et al. [34], but was computationally more intensive. We divided the data randomly 10,000 times into training and test sets of approximately equal size. We then subjected the 10,000 random training sets to logistic regression with forward selection on all 37 mass peaks to see how often each peak would be chosen for incorporation into a

multivariate classification model. We found that every peak was chosen at least once (the minimum was 12 times), that no peak was chosen 100% of the time (the maximum was 92.79%), and that the frequency of choice dropped sharply from the maximum (Figure 2). These findings demonstrate that the result of training a classification model under an automated variable selection procedure can be markedly sensitive to the particulars of a training-*versus*-test set division. In an effort to avoid such allocation artifacts, we chose two promising peak patterns to compare against forward selection for their average performance as multivariate logistic regression classifiers. The peaks at $m/z = 3966.8, 3983.1, \text{ and } 8951.7$ were common to both patterns.

For each pattern, we trained its classifier on the random training sets, tested the results on the corresponding random test sets, and estimated the expected number of correct classifications per test set for comparison. When the results for all 10,000 random test sets were plotted (Figure 3), it was clear that both patterns outperformed the set of models trained under forward selection. One pattern in particular, containing the three mass peaks above plus one other, did extremely well. From test set results, we also computed the average probability of being classified as cancer when in a test set for each patient under each peak pattern. Because this average probability per patient is based on that patient's appearance in about half of the 10,000 random test sets, it

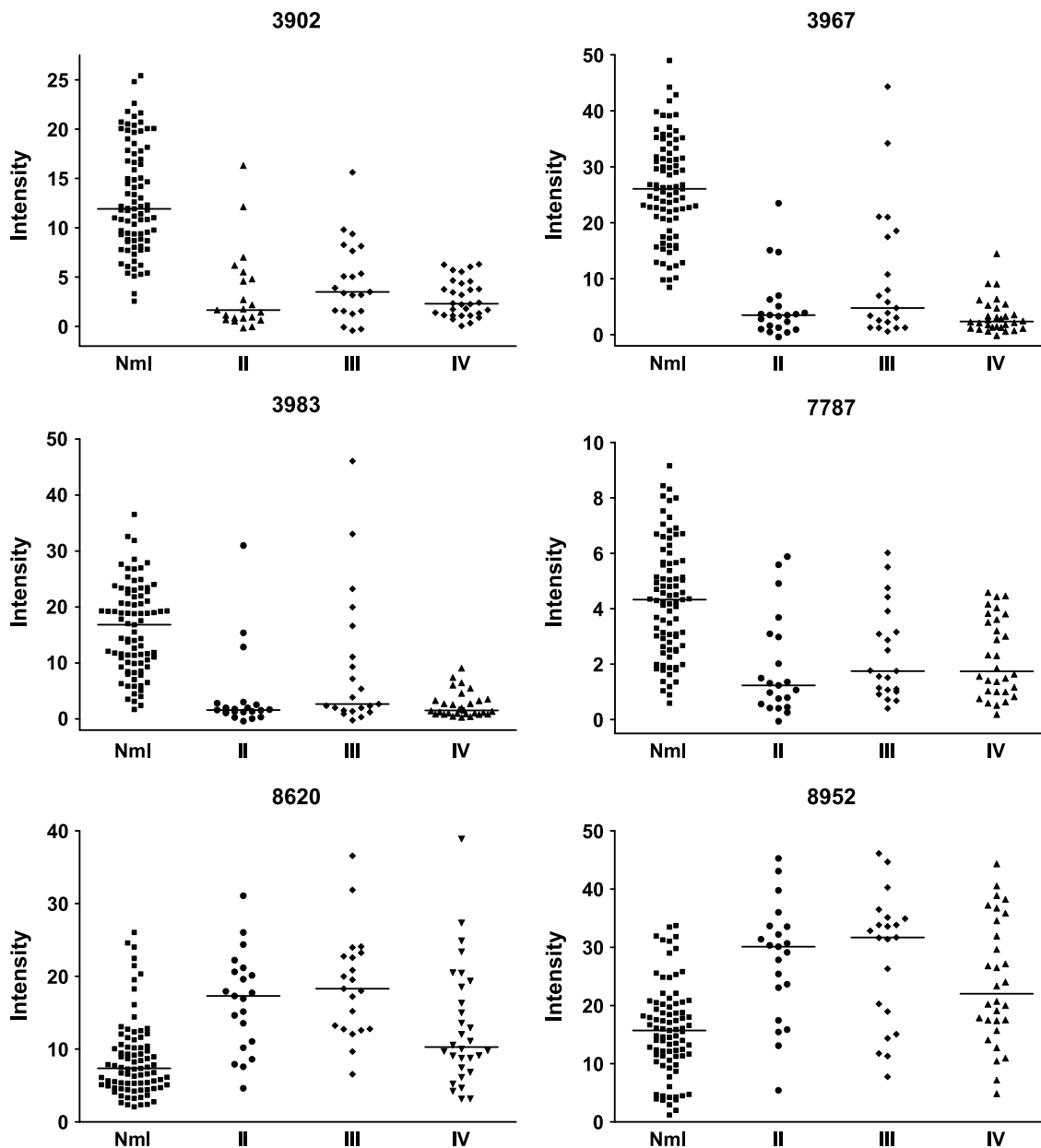


Figure 6. Group comparisons. Scatter plots of peak intensities from individual spectra of control serum (Nml) or sera from stage II (including one stage IB patient), III, and IV Pca patients for the eight peaks used in the logistic regression and CART classification models. Horizontal bar represents median peak intensities for each group.

can be interpreted as an estimate of the patient's fractional class membership that is independent of the particulars of training-*versus*-test set division, but still sensitive to the peak pattern and/or selection procedure used in the classifier. That interpretation should carryover to ROC curves constructed from such average probabilities (Figure 4). These considerations add to our confidence that the peaks we have identified have a classification proficiency that is biologically real and not an artifact of the choice of samples for the training set.

After cross-validation of the models, the spectra produced from the remaining 34 serum samples (16 PCa and 18 controls) were used as an independent test set to validate the two classification models. Both models correctly classified all of the PCa samples (100% sensitivity). The specificities determined for the two classification schemes were 93.5% for the CART model and 100% for the logistic regression model (Table 2). Included in the "normal" serum samples were seven patients with histories of other malignancies; thus, these models were able to correctly distinguish changes in serum protein profiles in PCa patients from other malignancies. Although the number and types of other malignancies included in this study are very limited, the results provide promise that this approach will be useful for specifically distinguishing PCa from other neoplasms based on a panel of serum biomarkers.

An examination of the mass peaks utilized by the two classification models reveals both increases and decreases in their median intensities when comparing the PCa serum proteome to that of the control serum (Table 1). This suggests that specific circulating serum proteins/peptides can be either elevated or diminished as a consequence of the disease state, and that these changes may be exploited to detect the presence of the tumor. Because the blood proteome perfuses the diseased organ, proteins abnormally shed by the tumor may add to the serum proteome, or enhanced proteolytic-degrading activities of the tumor may reduce specific serum protein levels [19]. A similar pattern of overexpression and underexpression of peptide/protein masses was observed by Adam et al. [30] in the serum protein fingerprint of their prostate cancer study. The identification of these complex proteomics patterns, reflecting both increases and decreases in specific serum proteins/peptides, underscores the power of this technology in the development of new diagnostic methods to detect early cancer.

Both classification models greatly improve the sensitivity and specificity of the current principal biomarker used for PCa, CA19-9. Due to the low prevalence of PCa, however, a much greater sensitivity is required before these types of approaches can be clinically useful for the early detection of PCa in an asymptomatic population. The predictive value of the mass peaks identified in this study requires further testing, including the examination of a larger panel of serum from patients with stage I disease, other malignancies, as well as benign diseases (including pancreatitis to assure that we are not merely identifying acute-phase proteins associated with inflammation in cancer). Thus, although our results

are very promising, they are not intended to be the final diagnostic paradigm.

In addition, protein discriminators may be added to these models after analysis of protein profiles produced from other serum fractions from these patients, as well as evaluation on other chip surfaces. For example, by combining the spectral data obtained from both weak cation exchange and copper metal affinity capture arrays, Bañez et al. [31] reported a significant enhancement in classification accuracy in their study of prostate cancer. Such studies are ongoing with our cohort. During the preparation of this report, a paper was recently published describing a similar diagnostic approach for PCa [34]. These authors also analyzed protein profiles from two ProteinChip surfaces (including the IMAC surface used in this study) as well as immunologic detection of CA19-9. Improvements in diagnostic accuracy were also noted for combinations between a particular marker panel derived from one SELDI surface and CA19-9; however, diagnostic panels were not reported for combinations of markers between chip surfaces or anion exchange fractions. Using a unified maximum separability algorithm to identify a discriminatory panel of markers to differentiate sera of PCa patients from control subjects, Koopmann et al. [34] achieved a maximum sensitivity and specificity of 78% and 97%, respectively, using two protein peaks identified from the WCX surface. The most discriminating peaks identified in fraction 1 from the anion exchange resin applied to the IMAC surface in their study, however, produced less effective results. Interestingly, using a similar fractionation strategy and ProteinChip IMAC surface, we have identified panels of four or five markers using either multivariate logistic regression or CART models that yielded 100% sensitivity and 93.5% to 100% specificity.

Besides the apparent difference in data analysis techniques utilized in these two studies, we attempted to identify other differences that may explain the disparity in our abilities to discriminate the PCa *versus* normal patient sera, particularly when comparing the same protein fractionation scheme and ProteinChip surface. Because both studies used approximately the same sample sizes to compare PCa *versus* healthy controls, we examined other differences between the patient populations. Both studies included patients with other pathologic conditions (e.g., inflammatory diseases) as controls to mute the identification of proteins involved merely in inflammatory responses, although our study did not include a patient population specifically for nonmalignant pancreatic diseases. However, this does not account for the discriminatory differences observed for PCa *versus* healthy individuals.

Although the mean age of PCa patients closely paralleled those used in the Koopmann study [34], we noted that the mean age of our control group was significantly lower. To examine the possible effects of using a control population with a lower mean age, we divided our control samples (mean age \pm SD) into a young group (41.7 ± 9.3) and an old group (67.3 ± 8.8) such that the mean age of the old group was similar to the cancer group. A comparison between the median peak intensities for the young and old

groups did not reveal any significant differences for any of the eight classification peaks (data not shown). Thus, the difference in mean age of our normal and cancer groups does not appear to have influenced our classification results. Other potential causes of the differences observed in these two studies include variations in the procedures used in collecting and handling serum specimens, differences in the alignment and calibration of the mass spectrometers, or lot variations in the preparation of the ProteinChip surfaces. It is intriguing, however, that with all the numerous factors that might explain the differences in our results, in both of our studies, one of the most discriminating markers identified from the IMAC profiles is the peak with m/z 3967 (reported as 3966.8 in our study). This peak was utilized as a discriminator in our multivariate logistic regression model (Table 2) and as the first decision node in the classification tree (Figure 5). In fact, this peak was incorporated into 93% of multivariate logistic regression models developed under forward selection (Figure 3), and was the major determinant for segregating 79% of the PCa specimens in the CART model (Figure 5, terminal node 1). Similarly, this peak was among the discriminators in the three-peak IMAC-Cu²⁺ panel described by Koopmann et al. [34]. With such a similar finding, it is tempting to speculate that we have both independently identified the same serum protein in our two studies. Thus, rather than highlighting the differences obtained in these studies using this technology, this observation strengthens the notion that SELDI protein profiling is a robust, reproducible procedure. The other two discriminators in their three-peak panel (3885 and 8929) are within approximately 17 and 23 Da of peaks we have identified (3902.2 and 8951.7); however, these would represent rather large deviations by mass spectrometry. In view of the 3967 peaks with essentially the same m/z in these two reports, it is questionable whether these peaks represent the same protein. These results, however, thus support the need to identify these proteins to substantiate these evocative findings.

The logistic regression classification equation (Table 2) and classification tree (Figure 5) described in this report should provide an easy means for validating future SELDI analyses performed in any laboratory using the same ProteinChip surface. After further validation and refinements to the classification models, proteomics fingerprints may directly aid in the diagnosis of PCa and other malignancies and, in themselves, constitute a valuable resource. The identification of the protein components of these fingerprints, however, will also provide important insights into the microenvironment of tumors and provide a better understanding of the processes involved in tumor development and growth.

Acknowledgements

We thank Jerry Malott, William Woodell, Conrad Browning, Bert Johnson, and the staff of the ACRC Outpatient Center Laboratory for help in collecting blood samples.

References

- [1] Ulrich CD II (2000). Growth factors, receptors, and molecular alterations in pancreatic cancer. *Med Clin North Am* **84**, 697–705.
- [2] Sakorafas GH, Tsiotou AG, and Tsiotos GG (2000). Molecular biology of pancreatic cancer; oncogenes, tumour suppressor genes, growth factors, and their receptors from a clinical perspective. *Cancer Treat Rev* **26**, 29–52.
- [3] Mack TM, Yu MC, Hanisch R, and Henderson BE (1986). Pancreas cancer and smoking, beverage consumption, and past medical history. *J Natl Cancer Inst* **76**, 49–60.
- [4] Lynch HT, Fitzsimmons ML, Smyrk TC, Lanspa SJ, Watson P, McClellan J, and Lynch JF (1990). Familial pancreatic cancer: clinicopathologic study of 18 nuclear families. *Am J Gastroenterol* **85**, 54–60.
- [5] Pietri F and Clavel F (1991). Occupational exposure and cancer of the pancreas: a review. *Br J Ind Med* **48**, 583–587.
- [6] What You Need to Know About Cancer of the Pancreas (2001). National Cancer Institute, Bethesda, MD.
- [7] Hawes RH, Xiong Q, Waxman I, Chang KJ, Evans DB, and Abbruzzese JL (2000). A multispecialty approach to the diagnosis and management of pancreatic cancer. *Am J Gastroenterol* **95**, 17–31.
- [8] Murr MM, Sarr MG, Oishi AJ, and van Heerden JA (1994). Pancreatic cancer. *CA Cancer J Clin* **44**, 304–318.
- [9] Rosewicz S and Wiedenmann B (1997). Pancreatic carcinoma. *Lancet* **349**, 485–489.
- [10] Moossa AR and Gamagami RA (1995). Diagnosis and staging of pancreatic neoplasms. *Surg Clin North Am* **75**, 871–890.
- [11] Posner MR and Mayer RJ (1994). The use of serological tumor markers in gastrointestinal malignancies. *Hematol/Oncol Clin North Am* **8**, 533–553.
- [12] Gullo L (1994). CA 19-9: the Italian experience. *Pancreas* **9**, 717–719.
- [13] Yasue M, Sakamoto J, Teramukai S, Morimoto T, Yasui K, Kuno N, Kurimoto K, and Ohashi Y (1994). Prognostic values of preoperative and postoperative CEA and CA19.9 levels in pancreatic cancer. *Pancreas* **9**, 735–740.
- [14] Satake K and Takeuchi T (1994). Comparison of CA 19-9 with other tumor markers in the diagnosis of cancer of the pancreas. *Pancreas* **9**, 720–724.
- [15] Satake K, Takeuchi T, Homma T, and Ozaki H (1994). CA19-9 as a screening and diagnostic tool in symptomatic patients: the Japanese experience. *Pancreas* **9**, 703–706.
- [16] Steinberg WM, Gelfand R, Anderson KK, Glenn J, Kurtzman SH, Sindelar WF, and Toskes PD (1986). Comparison of the sensitivity and specificity of the CA19-9 carcinoembryonic antigen assays in detecting cancer of the pancreas. *Gastroenterology* **90**, 343–349.
- [17] Steinberg W (1990). The clinical utility of the CA 19-9 tumor-associated antigen. *Am J Gastroenterol* **85**, 350–355.
- [18] Hasegawa T, Isobe K, Tsuchiya Y, Oikawa S, Nakazato H, Nakashima I, and Shimokata K (1993). Nonspecific crossreacting antigen (NCA) is a major member of the carcinoembryonic antigen (CEA)-related gene family expressed in lung cancer. *Br J Cancer* **67**, 58–65.
- [19] Wulffhulle JD, Liotta LA, and Petricoin EF III (2003). Proteomic applications for the early detection of cancer. *Nat Rev Cancer* **3**, 267–275.
- [20] Srinivas PR, Srivastava S, Hanash S, and Wright JR (2001). Proteomics in early detection of cancer. *Clin Chem* **47**, 1901–1911.
- [21] Negm RS, Verma M, and Srivastava S (2002). The promise of biomarkers in cancer screening and detection. *Trends Mol Med* **8**, 288–293.
- [22] Chambers G, Lawrie L, Cash P, and Murray GI (2000). Proteomics: a new approach to the study of disease. *J Pathol* **192**, 280–288.
- [23] Merchant M and Weinberger SR (2000). Recent advancements in surface-enhanced laser desorption/ionization time of flight-mass spectrometry. *Electrophoresis* **21**, 1164–1167.
- [24] Issaq HJ, Veenstra TD, Conrads TP, and Felschow D (2002). The SELDI-TOF MS approach to proteomics: protein profiling and biomarker identification. *Biochem Biophys Res Commun* **292**, 587–592.
- [25] Watkins B, Szaro R, Ball S, Knubovets T, Briggman J, Hlavaty JJ, Kusnitz F, Stieg A, and Wu Y-I (2001). Detection of early-stage cancer by serum protein analysis. *Am Lab*, 32–36.
- [26] Li J, Zhang Z, Rosenzweig J, Wang YY, and Chan DW (2002). Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* **48**, 1296–1304.
- [27] Poon TCW, Yip T-T, Chan ATC, Yip C, Yip V, Mok TSK, Lee CCY, Ho TWT, Leung SKW, and Johnson PJ (2003). Comprehensive proteomic profiling identifies serum proteomic signatures for detection of hepatocellular carcinoma and its subtypes. *Clin Chem* **49**, 752–760.
- [28] Petricoin EF III, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, Mills GB, Simone C, Fishman DA, Kohn EC, and Liotta LA (2002).

- Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**, 572–577.
- [29] Rai AJ, Zhang Z, Rosenweig J, Shih T, Pham ET, Fung LJ, and Sokoll DW (2002). Proteomic approaches to tumor marker discovery: identification of biomarkers for ovarian cancer. *Arch Pathol Lab Med* **126**, 1518–1526.
- [30] Adam B-L, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, Semmes OJ, Schellhammer Y, Yasui Z, and Feng GL (2002). Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* **62**, 3609–3614.
- [31] Bañez LL, Prasanna P, Sun L, Ali A, Zou Z, Adam B-L, McLeod DG, and Srivastava S (2003). Diagnostic potential of serum proteomic patterns in prostate cancer. *J Urol* **170**, 442–446.
- [32] Westfall PH and Young SS (1993). Resampling-Based Multiple Testing: Examples and Methods for p -Value Adjustment. John Wiley and Sons, Inc., New York, NY.
- [33] Breiman L, Friedman J, Olshen RA, and Stone CJ (1984). Classification and Regression Trees. Wadsworth International Group, Belmont, CA.
- [34] Koopmann J, Zhang Z, White N, Rosenzweig J, Fedarko N, Jagannath S, Canto MI, Yeo CJ, Chan DW, and Goggins M (2004). Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. *Clin Cancer Res* **10**, 860–868.

ERRATUM

Chromosomal Alterations during Lymphatic and Liver Metastasis Formation of Colorectal Cancer

Thomas Knösel, Karsten Schlüns, Ulrike Stein, Holger Schwabe, Peter Michael Schlag, Manfred Dietel and Iver Petersen

The image in Figure 1A of the article by Knösel et al. (*Neoplasia*, Vol. 6, No. 1, January/February 2004, pp. 23–28) is incorrect. It shows the CGH results of a subgroup of only 54 colorectal carcinomas and not the entire collective of 63 tumors as mentioned in the text. The larger collective shows slightly different frequencies of chromosomal imbalances for single chromosomes (e.g., a maximal 90% incidence of gains on chromosome 20q). The correct image as well as the histograms and CGH profiles of all individual tumors are accessible at the Charité CGH online tumor database at <http://amba.charité.de/cgh>.