[CANCER RESEARCH 63, 6971–6983, October 15, 2003]

Proteomic Profiling of Urinary Proteins in Renal Cancer by Surface Enhanced Laser Desorption Ionization and Neural-Network Analysis: Identification of Key Issues Affecting Potential Clinical Utility[1]

Mark A. Rogers, Paul Clarke, Jason Noble, Nicholas P. Munro, Alan Paul, Peter J. Selby, and Rosamonde E. Banks[2]

Cancer Research UK Clinical Centre [M. A. R., P. C., N. P. M., P. J. S., R. E. B.] and Department of Urology [A. P.], St. James's University Hospital, Leeds LS9 7TF, and School of Computing, University of Leeds, Leeds LS2 9JT [J. N.], United Kingdom

ABSTRACT

Recent advances in proteomic profiling technologies, such as surface enhanced laser desorption ionization mass spectrometry, have allowed preliminary profiling and identification of tumor markers in biological fluids in several cancer types and establishment of clinically useful diagnostic computational models. There are currently no routinely used circulating tumor markers for renal cancer, which is often detected incidentally and is frequently advanced at the time of presentation with over half of patients having local or distant tumor spread. We have investigated the clinical utility of surface enhanced laser desorption ionization profiling of urine samples in conjunction with neural-network analysis to either detect renal cancer or to identify proteins of potential use as markers, using samples from a total of 218 individuals, and examined critical technical factors affecting the potential utility of this approach.

Samples from patients before undergoing nephrectomy for clear cell renal cell carcinoma (RCC; $n = 48$), normal volunteers ($n = 38$), and outpatients attending with benign diseases of the urogenital tract ($n = 20$) were used to successfully train neural-network models based on either presence/absence of peaks or peak intensity values, resulting in sensitivity and specificity values of 98.3–100%. Using an initial "blind" group of samples from 12 patients with RCC, 11 healthy controls, and 9 patients with benign diseases to test the models, sensitivities and specificities of 81.8–83.3% were achieved. The robustness of the approach was subsequently evaluated with a group of 80 samples analyzed "blind" 10 months later, (36 patients with RCC, 31 healthy volunteers, and 13 patients with benign urological conditions). However, sensitivities and specificities declined markedly, ranging from 41.0% to 76.6%. Possible contributing factors including sample stability, changing laser performance, and chip variability were examined, which may be important for the long-term robustness of such approaches, and this study highlights the need for rigorous evaluation of such factors in future studies.

INTRODUCTION

There is currently considerable interest in using higher-throughput and multiple marker assays, both genomic and proteomic, to identify and evaluate new biomarkers for cancer (1). No routinely used tumor marker exists for renal cancer, which accounts for 2–3% of adult malignancies and causes up to 95,000 deaths/year worldwide (2). Often relatively asymptomatic, over half of all cases present with either local or distant metastases. Prognosis is mainly related to stage, with a 5-year survival rate in excess of 90% in stage I disease, but only 2–32% for stage IV (3, 4). There is clearly a need for clinically useful biomarkers to allow earlier detection, stratification/prioritization of investigation in symptomatic individuals, and the identification of patients at greater risk of progression who may benefit from adjuvant intervention (5).

In urological cancers, urine represents a particularly useful fluid in which to examine tumor markers because of its enhanced potential to contain higher concentrations of directly released tumor-derived products, and additionally its collection is noninvasive. The analysis of urine offers multiple technical challenges such as the several hundred-fold to thousand-fold more dilute protein concentration compared with serum, the presence of proteases, the variability of protein concentration because of hydration state, and the potential for contamination from other sources such as seminal or vaginal fluids. Systematic examination of urine composition by high-resolution two-dimensional PAGE and more recently by liquid chromatography/tandem mass spectrometry approaches has identified several hundred proteins, including intact and cleaved forms of plasma proteins such as retinol binding protein, transferrin, albumin and β_2-microglobulin, kidney or urogenital tract-derived proteins such as erythropoietin, urokallikrein, epidermal growth factor, E-cadherin, basement membrane antigens and Tamm-Horsfall protein, and viral or bacterial proteins (6–12). Tumor-associated proteins in urine include PCA-1 forms in prostate cancer (13), psoriasin (S100A7) in bladder squamous cell carcinoma (14, 15), tumor-associated trypsin inhibitor in many cancers (16), and several bladder transitional cell carcinoma markers including NMP-22,[3] BTA, and fibrin-fibrinogen degradation products (17, 18). Few urinary markers have been described in renal cancer, with NMP-22 (19, 20) and β-glucuronidase (21) being of potential interest.

A more recently developed complementary proteomic technology is SELDI, which is particularly biased toward investigation of molecules $<$20 kDa and has a sensitivity in the femtomole region or less (22). SELDI uses chip-based protein sample arrays of differing chemical chromatographic surfaces to selectively bind those proteins in a sample with specific chemical properties, for example hydrophobic, cationic, anionic, or metal binding molecules, before generating mass/charge profiles of the applied sample. Several examples now exist where SELDI has been used to identify cancer-specific protein "fingerprints" in tissue samples (23–26), quantitatively and qualitatively characterize existing known tumor markers in prostate cancer (27–29), to identify potential new markers such as hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein I in pancreatic ductal adenocarcinoma (30), and several potential urinary biomarkers in bladder cancer (29). Considerable potential has also been shown in combining this approach with the use of various computational models that, irrespective of knowledge of peak identity, can generate diagnostic predictions based on peak profiles, with promising results in studies using serum samples from patients with prostate cancer and breast cancer (31–34). In ovarian cancer, SELDI profiles of serum samples subjected to an iterative searching algorithm resulted in a model that identified patients with ovarian cancer, including patients with stage I diseases, from patients with benign conditions or healthy controls with a sensitivity of 100% and specificity of 95% (35).

Received 4/7/03; revised 6/30/03; accepted 7/24/03.

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 Cancer Research United Kingdom.

[2] To whom requests for reprints should be addressed, at Cancer Research UK Clinical Centre, St. James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom. Phone: 44-113-2064927; Fax: 44-113-2429886; E-mail: r.banks@leeds.ac.uk.

[3] The abbreviations used are: NMP, nuclear matrix protein; BTA, bladder tumor antigen; SELDI, surface enhanced laser desorption/ionization time of flight mass spectrometry; RCC, renal cell carcinoma; CV, coefficient of variation.

Although such results are extremely impressive, their validation and utility over a longer time period have yet to be explored.

We have examined the clinical utility of the combination of neural-network modeling and SELDI profiling using urine samples from patients with renal cancer, initially examining the ability of such models to discriminate between normal and malignant disease, and also the potential to differentiate from various benign urological conditions. As an additional examination of the robustness of the technique, the predictive accuracy of such models when applied to similar data collected 10 months later has also been explored.

MATERIALS AND METHODS

Materials. Weak cation exchange (WCX2) protein chips and other SELDI-related materials were obtained from Ciphergen (Fremont, CA). Other materials were sourced as follows: Complete protease inhibitor mixture tablets and NP40 were from Roche (Lewes, United Kingdom), Multistix from Bayer (Newbury, United Kingdom), Falcon nylon filters from Becton Dickinson (Crawley, United Kingdom), PBS tablets from Oxoid (Basingstoke, United Kingdom), and trifluoroacetic acid from Pierce (Tattenhall, United Kingdom). All of the other general chemicals were purchased from BDH (Poole, United Kingdom). Milli-Q water was used throughout.

Sample Collection. Voided urine samples were obtained from a total of 218 individuals. For the main body of the study ($n = 138$; see Table 1), samples were obtained from patients before undergoing nephrectomy for later histologically confirmed clear cell RCC ($n = 60$), normal volunteers ($n = 49$), and outpatients attending with benign diseases of the urogenital tract ($n = 29$). For the subsequent long-term testing of the technique, urine samples from an additional 80 subjects were used comprising 36 patients with RCC, 31 healthy volunteers, and 13 patients with benign urological conditions (Table 2). This study was approved by the Local Research Ethics Committee, and all of the participating patients gave informed consent. A sample of the urine obtained was assayed by standard automated method for protein and creatinine, using the Synchron LX-20. Within 15 min of voiding, each urine sample (50–100 ml) was placed on ice and a Complete protease inhibitor mixture tablet added. Hematuria was assessed using Multistix, and the pH was measured and adjusted to 7 (by addition of 1 M NaOH or 1 M HCl as appropriate). The adjustment of the pH to neutral before freezing minimizes precipitation during storage (36). Samples were then sieved through a 100-μm nylon filter to remove cellular debris and aggregates, centrifuged for 10 min at 800 $\times$ *g* at 4°C, and aliquotted and stored at −80°C. The period of storage for samples used in this study ranged from 2 weeks to 2 years.

Matrix Preparation and SELDI Calibration. Calibration of the SELDI system was carried out before each analytical session. Sinapinic acid matrix was prepared freshly on the day of use by sequential addition of 125 μl of acetonitrile and 125 μl of 1% trifluoroacetic acid to an aliquot of matrix. This was vortexed and stored in the dark. Before use the matrix was microfuged at 13,000 rpm for 1 min. For calibration purposes, an H4 chip spot was loaded with 2 μl of matrix containing 400 fmol of bovine ubiquitin (M_r 8,564.8), 400 fmol of bovine superoxide dismutase (M_r 15,591.4), and 400 fmol of bovine β lactaglobulin A (M_r 18,363.3) calibrants. After air-drying, the calibrant mixture was analyzed on the SELDI system using parameters of a high mass acquisition of 40 kDa, optimum mass range 3–20 kDa, laser intensity of 210, sensitivity of 10, and collecting 50 transients across the spot surface. Calibration was performed using single and double charged peaks for each calibrant.

Table 1 *Details of the patients and control groups for samples included in the main study groups for establishment and testing of the neural network*

Group	Sex	Mean age (range)	Clinical details	
			Subgroup	Number
Normal healthy volunteers ($n = 49$)	30 male 19 female	54 (23–95)	—	—
Benign urological diseases ($n = 29$)	21 male 8 female	63 (29–88)	BPH	11
			Stones	5
			Outflow dysfunction	6
			UTI	7
RCC ($n = 60$)	37 male 23 female	64 (42–85)	Stage T1	23
			T2	8
			T3	26
			T4	3
			Grade 1	4
			G2	14
			G3	30
			G4	12

Table 2 *Details of the patient groups used for the long-term "blind" test set*

Group	Sex	Mean age (range)	Clinical details	
			Subgroup	Number
Normal healthy volunteers ($n = 31$)	22 male 9 female	63 (37–81)	—	—
RCC ($n = 36$)	27 male 9 female	56 (23–75)	Stage T1	13
			T2	5
			T3	16
			T4	2
			Grade 1	3
			G2	9
			G3	15
			G4	9
Benign urological illnesses ($n = 13$)	7 male 6 female	58 (34–86)	BPH	2
			Stones	4
			Cysts/oncocytoma/adenoma	5
			UTI	2

Preparation of Protein Chips for SELDI Analysis. Samples from the groups in Table 2 were analyzed by SELDI 8–10 months after the initial analysis and establishment of the neural net models using those in Table 1. On the basis of optimization studies as described in the following sections, the following protocol was adopted for the standard analysis of all of the urine samples in the study. Urine samples were thawed and microfuged (10,000 $\times$ *g*) for 10 min. Samples were then diluted to a protein concentration of 50 μg/ml, spiked with additional Complete protease inhibitor mixture (equivalent to 1 mini protease tablet/100 ml of diluted sample), and placed on ice.

After initial pilot experiments, the chip type selected for this study was the weak cation exchange (WCX2) chip. Protein chips (8 spot) were prepared by pretreatment with 5 μl of 10 mM HCl/spot (2 $\times$ 5 min) followed by three rapid washes, each with 5 μl of water. The urine samples prepared above were diluted 1:1 with 2$\times$ binding buffer [40 mM ammonium acetate/0.2% v/v NP40 (pH 6.5)] containing 4 femtomol/μl bovine cytochrome C (M_r 12,230.9) as an internal calibrant and 50 μl of diluted sample applied/spot (1.25 μg total protein) using a bioprocessor. After incubation at room temperature with shaking at 200 rpm for 30 min, samples were removed and the chips washed five times (100 μl/spot for 1 min each wash) with 1$\times$ binding buffer followed by two washes as above but with water, and allowed to air-dry. To each spot, 0.35 μl of sinapinic acid matrix solution was added, allowed to dry, and an additional 0.35 μl of matrix was added. After air drying, the chips were analyzed on the SELDI system with PBS-II software using a protocol with the following parameters: acquisition up to 100 kDa, optimum mass range 3–20kDa, laser intensity of 210 with two warming shots not collected, sensitivity of 10, and collection of 50 transients across the spot surface.

Determination of Optimum Sample Load. To examine the effects of total protein amount on the SELDI profiles, 6 urine samples (3 normal and 3 RCC) were prepared as described previously with final protein concentrations of 50 μg/ml, diluted 1:1 with 2$\times$ binding buffer, and final volumes varying from 5 μl (125 ng total protein applied directly to the chip) to 250 μl (6.25 μg total protein applied using the bioprocessor) were loaded on to WCX2 chips and subjected to SELDI analysis as described above. Similar experiments but examining the effect of various protein concentrations of the diluted samples from 25 μg/ml to 6.25 μg/ml were also carried out. The effects of salt and urea on profiles were also examined by adding appropriate amounts of stock solutions of urea and NaCl to produce elevations of 20 and 100 mM for each.

Assessment of Sample Stability. To evaluate stability during experimental processing before SELDI, several experiments were carried out on a total of 8 urine samples (processed as described above, *i.e.*, with the addition of protease inhibitor before freezing), where samples were thawed and either no additional protease inhibitor was added or protease inhibitor equivalent to 1 mini protease inhibitor tablet/100 ml was added. Samples were then analyzed by SELDI immediately and after being left on ice for up to 2 h. The effects of the additional protease inhibitor *per se* on peak profiles were examined by com-

parison of peak intensities of paired samples (plus and minus inhibitor addition) at time zero. Sample degradation was examined by analyzing changes in peak intensities across the 2-h period after thawing given the addition of additional protease inhibitor and changes in peak intensities across this same time period without the addition of additional inhibitor.

Assessment of Reproducibility. To assess intrachip variability, 50 μl of 2 urine samples, 1 normal and 1 RCC, were each loaded onto 4 spots of a WCX2 chip as described previously with a total protein load per spot of 1.25 μg. To assess interchip variability, the same samples were assayed using three different chips. Profiles were normalized for mass to the internal calibrant (single and double-charged peaks), and reproducibility of the mass determination and of the peak intensities were calculated for five selected peaks in each sample ranging from 3 to 20 kDa and of various signal intensity, exporting the peak information for additional analysis.

Data Processing. Data were processed and examined in two distinct ways. The calibration of all of the spectra was checked using the peaks of the internal calibrant at 12230.9 and 6115.4 and recalibrated where necessary. In the first approach, the Biomarker Wizard feature on the standard Ciphergen ProteinChip software version 3.0 was used to autodetect the peaks present in all of the 138 samples used in the initial training sets and blind sets. The spectra were examined in two separate regions of 2500–20,000 Da (the lower region being excluded because of matrix interference making this region unreliable) and 20,000–100,000 Da as different peak detection settings were found to be optimal for each because of different noise levels and peak resolution stringency. These were determined by iterative processing of spectra and monitoring the changes seen with modifications in the different parameters until optimal peak detection was achieved visually. For the lower molecular weight region, settings of 2 σ for the noise determination using the spectral mass region of interest only, a signal:noise ratio >5 was used for first pass with clusters being completed using signal:noise ratio of >3 for second pass, cluster mass setting of 0.2% (*i.e.* $\pm 0.1\%$), and a requirement to be in a minimum of 10% of samples was used. For the higher molecular mass region, noise was set at 3 σ and the cluster mass of 2% otherwise values were the same as for the low region. The Biomarker Wizard was then used to determine clusters significantly different among the normal control, benign, or RCC patient groups using standard nonparametric tests. Additionally, after peak detection manually using these settings but with no cluster analysis, raw data were exported to Microsoft Excel to determine the presence/absence of peaks with values of zero assigned to spectra for absent peaks and the number of spectra containing each peak determined.

In the second approach, the spectra generated by SELDI were processed before evaluation by neural-network analysis. This is described in detail below. In summary, the SELDI profile data from a subset (training set) consisting of 48 RCC samples, 38 controls, and 20 benign samples of the main study group of patients and controls (Table 1) were analyzed to locate spectral peaks that, considered alone, were valuable in distinguishing cancer patients from controls as described below. The results of this analysis were, in turn, used to generate multivariate models through the application of neural-network classification techniques. Two models using a control set composed of either healthy controls or healthy controls combined with benign controls were both tested. In addition, the effect of correcting for creatinine by scaling peak intensity values by a factor derived from the protein/creatinine index was examined, generating two additional models. A simpler method, taking the binary presence or absence of each peak, regardless of its intensity, was also examined. The success of each model in correctly classifying a "blind" test set of 12 RCC samples, 11 controls, and 9 benign samples for which the programmers were unaware of the diagnosis was then measured. As an additional test of the applicability of the technique and model over long-term use, an additional "blind" group of samples consisting of 36 RCC, 31 controls, and 13 benign samples (Table 2) was analyzed by SELDI over a period of 2–3 months 9–10 months later and against tested the neural-network models.

Peak Detection and Selection Procedure. The data output from the SELDI mass spectrometer consists of ~32,000 data points for each sample profiled with a high mass acquisition of 100,000 Da, each one associating a *m*/*z* with an amplitude. In the region between 2,500 and 20,000 Da where the majority of peaks were seen, there are ~10,000 data points for each spectrum.

To reduce the spectral data to a smaller number of key variables and enable the construction of valid predictive models, a peak detection algorithm was generated, taking into account the requirement to discriminate peaks from noise and the variance in peak width in different regions of the spectra. The algorithm examined each of the 10,000 data points in the 2,500–20,000 Da range to determine which of these could form the center of a peak. A peak was defined as a high intensity value with a given number of progressively lower intensity values immediately to both sides. A varying size window was passed over the data, with the number of data points required to form a peak varying between 2 points and 20 points on each side of the central data point under consideration. To overcome the minor fluctuations because of noise, the data were smoothed using a simple n-point moving average where n is equal to the window width.

The results of this peak detection for each of the 106 training spectra and the

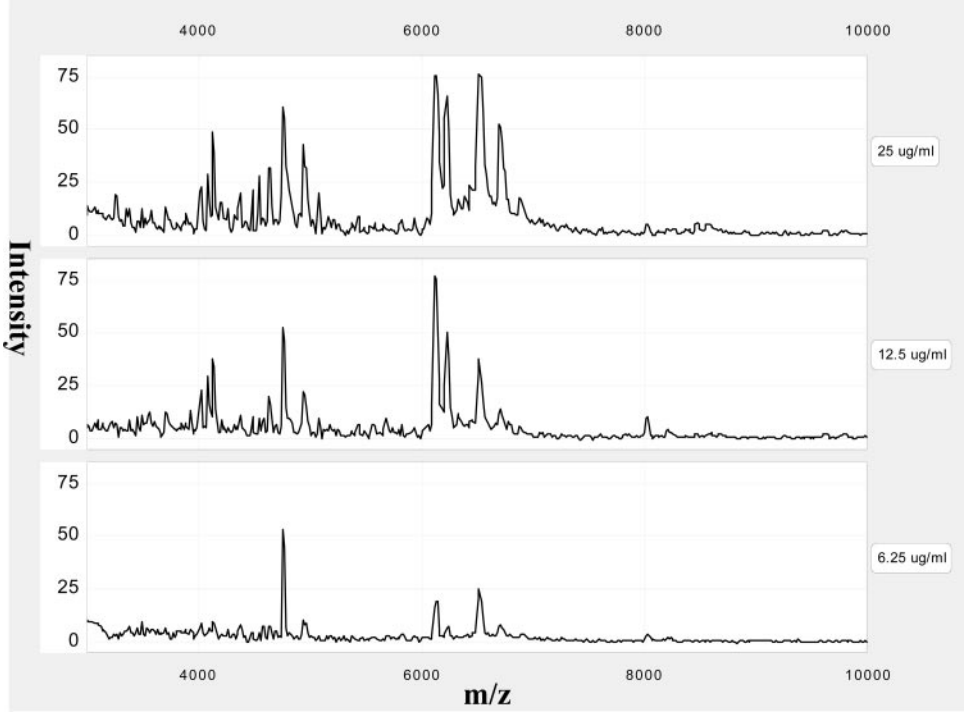

Fig. 1. SELDI spectra of a urine sample from an RCC patient illustrating the effect of sample protein concentration on the profiles obtained. WCX2 chips were loaded using a bioprocessor as described with 50 μl of diluted urine sample varying from 25 μg/ml to 6.25 μg/ml final concentration. Similar effects were also seen if total protein was kept constant by loading different volumes of the diluted urine samples to compensate.

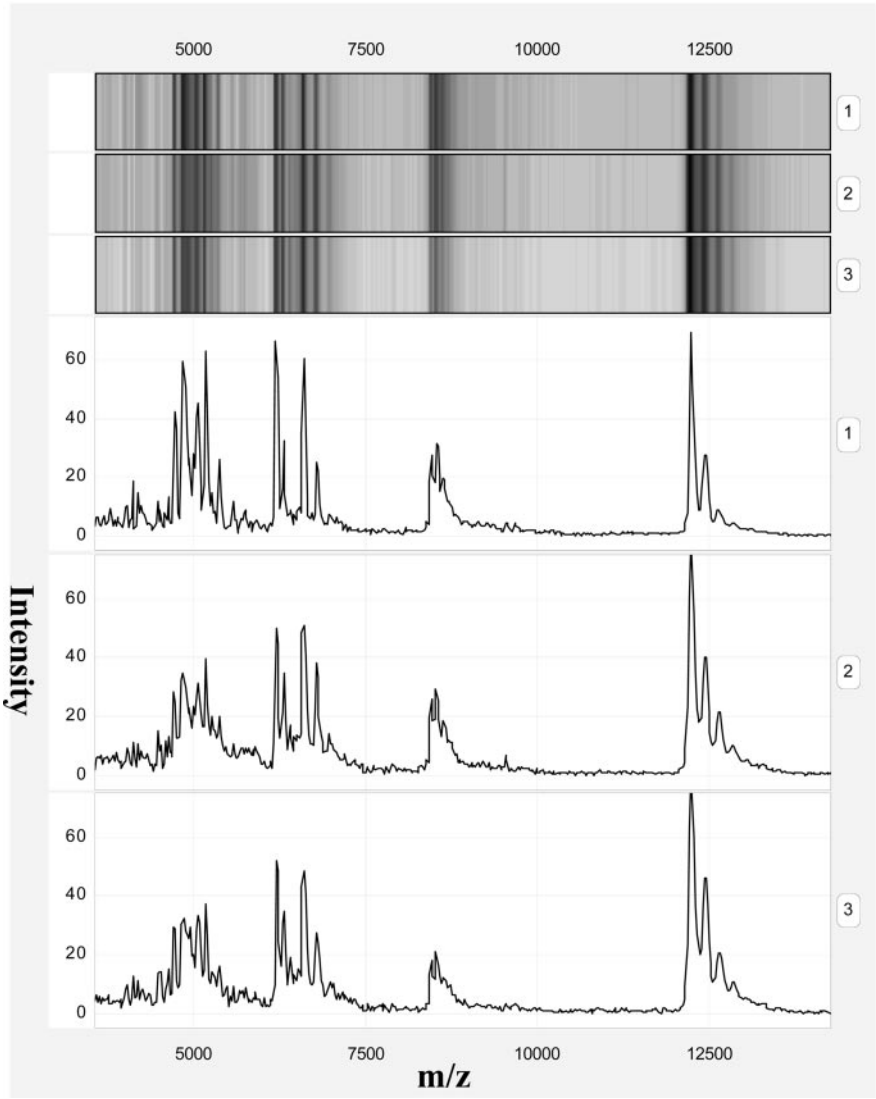

Fig. 2. *A*, SELDI analysis of a normal urine sample repeated on three separate experiments to illustrate the interchip consistency within the region up to 12 kDa where the majority of peaks are seen, and *B*, SELDI analysis of 2 fresh urine samples before addition of any protease inhibitor and after spiking with protease inhibitor mixture as per the standard sample processing protocol before freezing. The "gel-based" profiles are also shown as an illustration of the alternative data presentation mode.

32 blind spectra were combined to locate clusters of peaks with similar *m*/*z* values. This involved forming a frequency distribution of the 10,000 *m*/*z* values against the total number of peaks identified within a tolerance of 0.1% of each one. The peaks within this distribution formed the center points of clusters of peaks that could be considered the "same" peak: 368 "master" peaks were identified in this way.

The final stage of peak detection was to build an input pattern for each sample. These patterns consisted of one value for each of the 368 master peaks, namely the intensity of any peak in the sample within 0.1% of the master peak, or zero if no such peak was present. These were then scaled into the range of (0, 1) for each input. For the binary models, this was simplified to an input of one for present or zero for absent. Some statistical analysis was required to select which of the inputs would be included in the classification model. A 2×2 table was formed for each input showing the incidence of peaks against the incidence of cancer over the training set. From these, the χ^2 indicator for each input was calculated, and this sorted the inputs into order of significance.

Neural-Network Classification Procedures. Fully interconnected feed-forward neural-networks were set up with varying numbers of the most significant inputs, five hidden-layer neurons, and one output neuron. The hidden neurons mapped the sum of their inputs to their outputs using a standard sigmoid function ($1/1+e^{-x}$). A prediction of cancer corresponded to a high signal on the output neuron. All of the connection weights were randomly initialized in the range $(-1, +1)$. The network was then presented with the useful-peak data for each of the subjects, and trained using the back-propagation algorithm (37). Subjects were presented in random order, alternating between positive and negative classifications to prevent any bias forming because of their unequal numbers. Training was complete when the network correctly classified all of the input data or when 100 presentations of the data set had been completed, whichever came first. The learning rate and momentum term were set to 0.45 and 0.90, respectively.

The process of predictive model construction was conducted six times, once using only the 38 normal controls as comparison and a second time including 20 benign controls for a broader control group of 58 cases, with each, in turn, being subjected to creatinine correction and then simplification to binary input as described previously. The predictive accuracy of each neural-net model was assessed by using it to predict the status of the initial blind test group and subsequently the larger, later time point blind group (Table 2). Because of the stochastic nature of the training process (random initial weights; random order of presentation of subjects) there was a need for repeated training runs. Thus, when a particular neural-net model is spoken of below, we are, in fact, referring to the averaged outcome of 10 models run with identical parameters.

RESULTS

Preliminary work with urine showed WCX-2 chips to provide the most promising spectra to allow discrimination between these sample

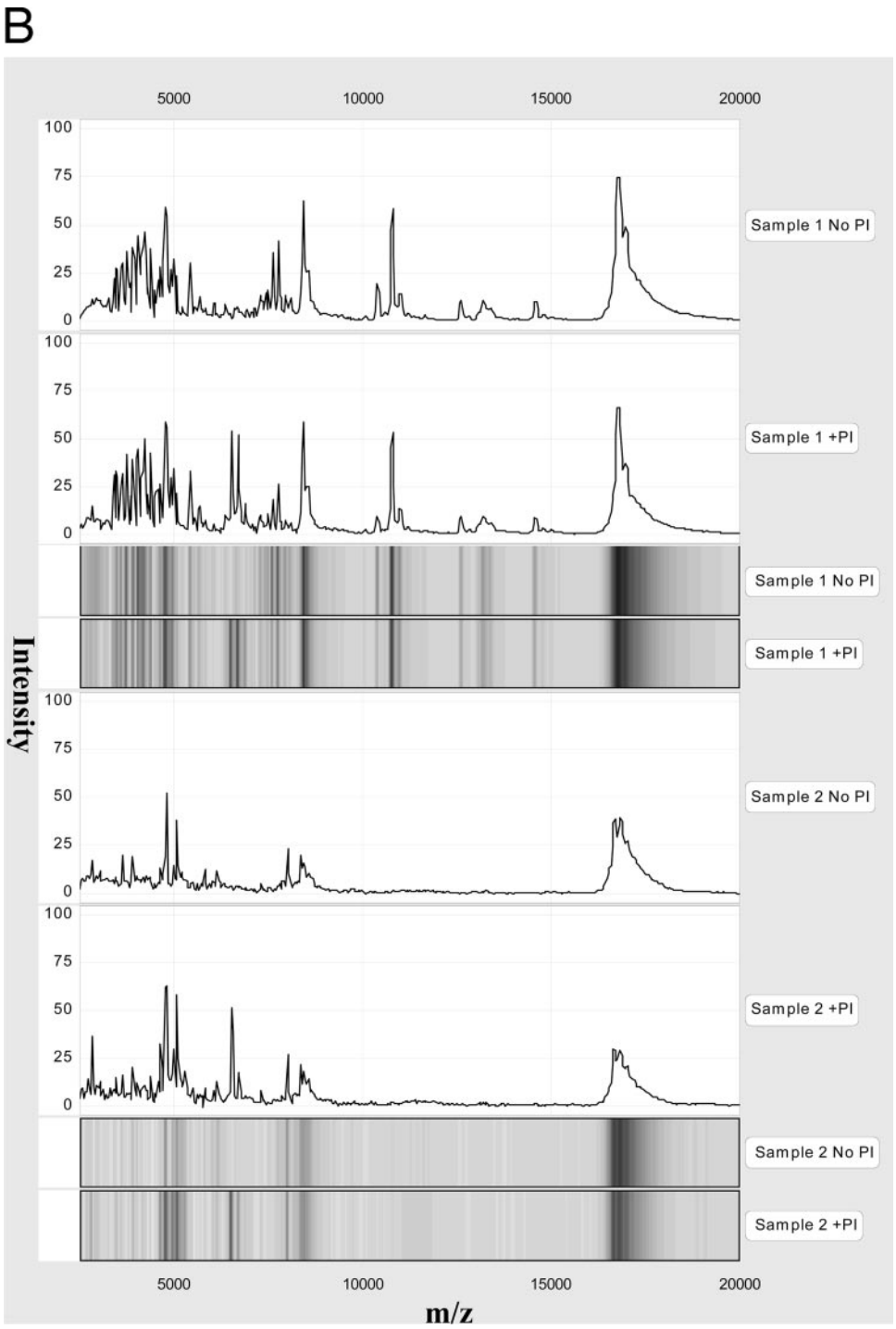

Fig. 2. *Continued.*

groups when processed by the method described (data not shown). Titration of the total protein loaded by varying the diluted urine volume loaded from 5 μl to 250 μl at a constant concentration of 0.025 g/liter protein showed marked differences in numbers of peaks detected between 5 μl and 50 μl loads, with 50 μl being maximal (*i.e.*, 25 μl of urine and 25 μl of 2$\times$ buffer, representing a protein load of 1.25 μg). Little difference was generally seen between 50 and 200 μl sample volumes (*i.e.*, 1.25–5 μg protein). However, the most marked effects were seen with sample concentration for all 4 of the samples examined where more dilute samples clearly provided inferior quality spectra, irrespective of whether they were normal or tumor urine samples (Fig. 1). Increasing sodium chloride or urea concentrations of the diluted samples by 20–100 mM had no effect on protein profiles obtained.

Analysis of five peaks ranging from 2,886.1 to 16,856.7 Da showed good reproducibility of mass determination with CVs for both inter- and intrachip assessment being $\leq$0.05% in all of the cases ($\leq$0.02% in 50% of cases) with the exception of two peaks in the RCC sample where CVs of 0.1% and 0.09% were found. As expected, the CVs for peak intensity were considerably greater, with selected peaks having mean peak intensities ranging from 4.6 to 44.6 and CVs 14.0% to 57.4% (average of 34.2% and values $<$30% in 50% of cases) with similar values for inter- and intrachip comparisons. Examples of the degree of reproducibility are shown in Fig. 2*A*.

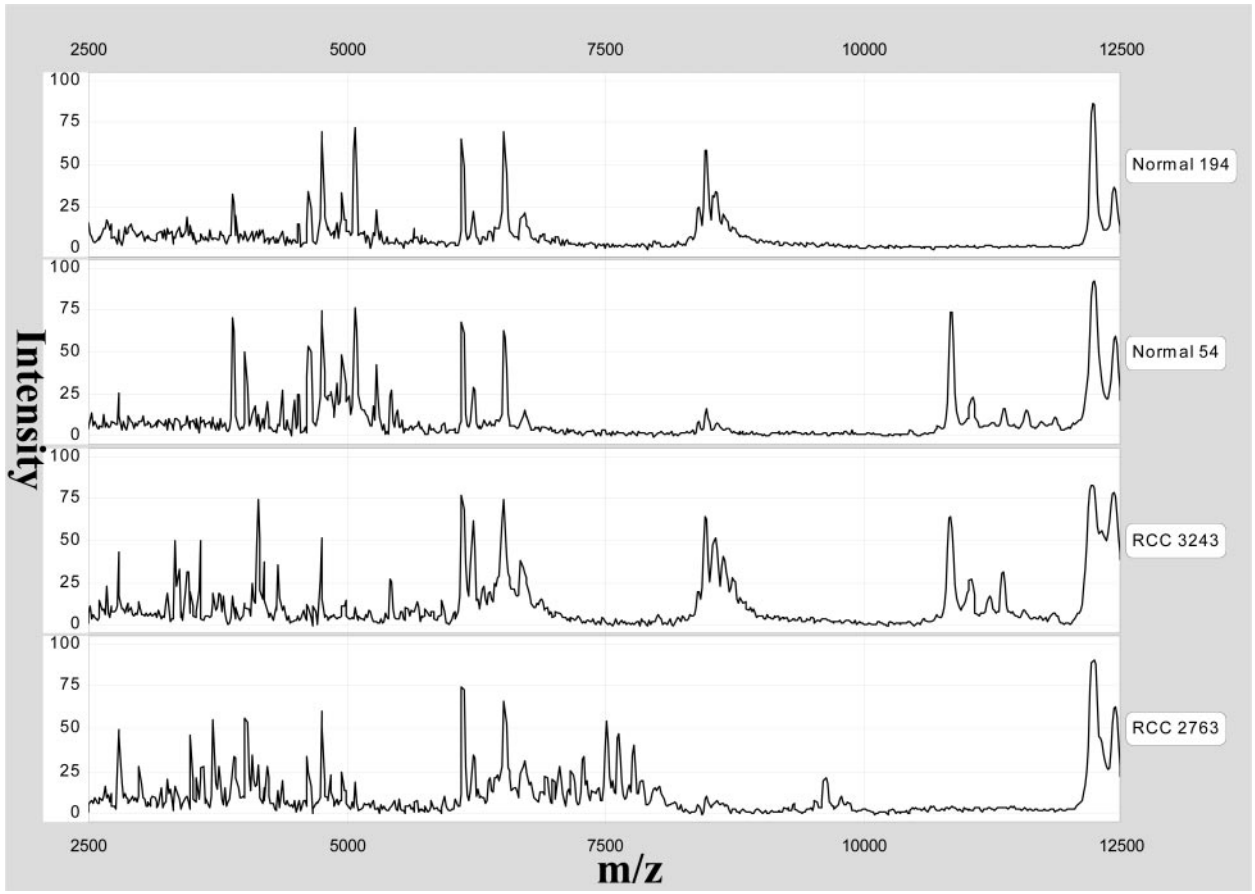

Fig. 3. Examples of SELDI WCX2 profiles of 2 normal and 2 RCC urine samples illustrating the major peaks and the interindividual heterogeneity seen. The peak of aprotinin resulting from the protease inhibitor addition can be seen at 6,504 Da as can the single and double-charged peaks of internal calibrant cytochrome C at 12,230.9 and 6,115.5, respectively.

The samples available for this study from our sample bank had been standardly processed, being placed on ice immediately after collection from the patient and protease inhibitors added to minimize degradation. From a gross comparison of 3 samples analyzed fresh and after short-term freezing, the freezing process *per se* did not appear to alter the profiles and comparison of 2 samples before and after addition of inhibitors showed no gross effect of the inhibitor mixture on the profile *per se*, which would preclude their use in this study (Fig. 2*B*). In early studies examining sample stability after thawing and while on ice before immediate application on to the SELDI chip, 2 samples that had been processed and stored frozen before analysis showed some evidence of degradation. For this reason and because some inhibitors are not stable for prolonged periods of storage, addition of further protease inhibitors after thawing for analysis by SELDI became routine practice. Analysis of the 8 samples in each group (*i.e.*, plus and minus additional inhibitor mixture) showed 67 peaks present in at least 1 sample, and the mean change in intensity over the 2-h period across 8 samples for each of these peaks was noted. The mean value of these 67 means was not significantly different from zero in either case (protease inhibitor added, $t_{66} = -1.24$, $P = 0.22$; no protease inhibitor added, $t_{66} = -0.59$, $P = 0.56$) indicating no evidence of systematic degradation or indeed amplification of peak intensities. For the matched comparisons of sample ± inhibitor mixture, there were 72 peaks present in at least 1 sample, giving 72 mean differences in peak intensity across the 8 subjects. The mean of these mean differences was again not significantly different from zero ($t_{71} = -1.36$, $P = 0.18$) confirming the lack of systematic effect of additional protease inhibitor on spectra *per se* other than the peak of aprotinin at 6504 Da.

From the manual analysis using the standard Ciphergen software to examine the spectral profiles and generate cluster reports in the range 2,500 to 20,000 Da, a total of 180 peaks were detected in the initial 138 samples used. Using the threshold of being present in at least 10% of spectra and the peak clustering parameters indicated, this figure was reduced to 30 peak clusters excluding those derived from the cytochrome C spike or aprotinin. Peaks identified as being consistently present in the majority of samples irrespective of sample grouping were present at 4,753.3, 6,715.6 and 16,939.0 Da. Sixteen peaks were identified as being significantly different between the normal and RCC groups with the most significant of the positive discriminators for RCC being 2,789.5 ($P = 0.0196$; 32 RCC *versus* 21 normal), 3,580.7 ($P < 0.0001$; 11 RCC *versus* 0 normal), and 4,136.3 Da ($P = 0.0113$; 19 RCC *versus* 9 normal). The peak at 3580.7 was the only peak found solely in the RCC group, although only present in 11 samples. Interestingly, this peak was biased toward patients with larger tumors with 3 of 3 patients with T4 tumors, 6 of 26 T3 samples, 2 of 8 T2 samples, and 0 of 23 T1 samples being positive. The most significant discriminatory peaks for normality were at 4,957.7 ($P < 0.0001$; 37 normal *versus* 26 RCC), 5,071.5 ($P < 0.0001$; 37 normal *versus* 19 RCC), 5,276.3 ($P = 0.0001$; 20 normal *versus* 7 RCC) and 16,793.3 Da ($P = 0.01$; 33 normal *versus* 17 RCC). Examples of some spectra illustrating these peaks are presented in Fig. 3.

The data collection and indeed calibration of spectra was optimized for peaks below 20,000 Da. However, the peak clustering above this size was also examined. Twenty peak clusters were identified, the majority of which were of intensity <1. The dominant peak present in >90% of samples was 23,882 Da. Rarely, small amounts of a peak

corresponding to approximately 66–68 kDa and, therefore, likely to be albumin were detected. Two peak clusters were found to be significantly increased between normal and RCC samples ($P = 0.0098$ for both), namely at 39,943.5 and 74,926.5 Da, the first of which apparently represented the doubly charged daughter peak of a protein of approximate molecular size 79–80 kDa, although present in relatively few samples and of very low intensity.

Neural-network analysis of the complex spectra obtained using SELDI initially appeared to be successful in that both the models using either presence/absence of peaks or the intensities of peaks were able to discriminate between RCC and healthy normal controls in the training sets with sensitivities and specificities of 98–100% (Tables 3 and 4). Using an initial set of 32 samples tested blind against the models, variable results were achieved, and in general the models incorporating peak intensities performed better than those solely examining presence and absence of peaks with sensitivities and specificities of >80% for all of the group comparisons (Table 3). The inclusion of benign samples in the control training set produced poorer results in the presence/absence model than training solely on healthy controls whether discriminating between RCC and healthy controls alone, or healthy controls and benign combined (Table 4). This was not the case for the intensity model where either training set produced similar results with sensitivities and specificities >80% (Tables 3 and 4). The peaks found to be most discriminatory and incorporated into the neural-network from all of the peaks detected in the samples significantly overlapped with those found by manual analysis to be significant, particularly those at 5071.5, 16793.3, 5276.3, 4537.5, and 4957.7 Da. In the benign group, BPH and benign kidney diseases were generally predicted well, whereas the urinary tract infection samples were all predicted as RCC.

However, the results achieved using a larger set of samples tested blind ($n = 80$) ~10 months later were disappointing. For the presence/absence models, although specificities were either little affected or indeed improved, sensitivities decreased markedly to <40% (Table 5). For the intensity-based model, sensitivity and specificity declined to 50–70%. Similar poor results were obtained when using the models trained on a control group including the benign samples (data not shown).

To investigate these results, the standard Ciphergen ProteinChip software was used to analyze the late blind results and the results from 6 additional samples (3 normal and 3 RCC), which had been run originally in the training set and which were also rerun at the time of the late blind set for quality control. Analysis of the blind set in a similar way to the original training set showed that although all of the peak clusters detected in the training samples were also present in the late blind set, the only peak to now show significant differences between the groups was 5071.5 Da ($P = 0.046$), although now being present in 28 of 34 (82.4%) normal and 22 of 39 (56.4%) RCC samples compared with 37 of 49 (75.5%) normal and 19 of 60 (31.7%) RCC samples in the original set. Similarly the peak at 3580.7 Da, which was originally only detected in 11 RCC samples, was now detected in 7 RCC samples but also in 4 normal samples. Of the 6 samples that were rerun, only 1 normal and 2 RCC samples were correctly predicted. On the basis of comparing signal intensity of routinely run protein mixtures at this time with those a year earlier, peak intensities were obviously lower. Similarly, visual inspection of the spectra showed that although qualitatively the presence of many similar peaks could be seen in both runs, the spectra in the late blind run contained peaks of generally lower intensity and less apparent complexity than those produced initially. Assuming a decline in laser and detector performance with age, an attempt to compensate for this by rerunning the late blind samples with various combinations of laser intensities and detector voltages failed to improve the results markedly with the neural-network model. This may have been partly because of the maximum achievable increase in laser intensity only being from 210 to 200, because at higher settings the matrix signal impacted considerably on the peaks in the lower regions of the spectra, but may also indicate other potential contributing factors.

A later analysis of chips from different batches revealed an additional possible contributory factor in that the late blind set of samples was analyzed gradually over a period of 3 months using several batches of chips but not those used for the initial analysis. Unfortunately none of the initial batches of chips was available for repeat examination with the exception of 4 spare spots on a chip batch from a similar time to those used for establishment of the initial model, although now significantly past its validation date (AB058). It was also possible to examine 2 samples on a limited number of spare sample chip spots from several batches of chips including some used in the late blind trial. It is quite evident that although in most cases the overall gross profiles are similar, there are both qualitative and quantitative variations between chip batches with several of the peaks found to be the most significant discriminators being most affected by the use of different chips. An example of the effects seen is shown in Fig. 4. Clearly some chips such as DU081 and AB058 (although 1 year past validation date) produced very similar profiles to those used originally, although particularly and not unexpectedly in the latter case, with reduced signal. However, using the identical sample solution, CN651 showed selective loss of some peaks including the cluster at 5 kDa, and DZ145 showed a more similar profile to the early results but markedly reduced signal across the range. These effects were reproducible in that alternative CN and DZ chips from the same batches showed the same differences in profile.

After the recent replacement of the laser and detector, 39 samples (19 RCC and 20 control) from the original training set used to establish the neural-network ~18 months earlier were rerun at a variety of laser intensities and tested blind against the original network. Laser performance was clearly vastly improved with the best results visually in terms of comparability between runs being obtained at a laser intensity of only 190 compared with 210 when the samples were originally used to generate the neural-network (Fig. 5). Indeed, comparison of data generated with the new laser at a setting of 210 was not performed, as the signal was so intense that saturation of

Table 3 *The results of the neural-network models based either on peak presence/absence or peak intensity values trained on RCC versus normal healthy controls*
The results are given for the training set and the initial blind set.

	Presence/absence model				Intensity model			
Sets	Sensitivity	Specificity	PPV[a]	NPV	Sensitivity	Specificity	PPV	NPV
Training								
RCC (48)[b] *vs.* healthy controls (38)	100%	100%	100%	100%	100%	100%	100%	100%
Initial blind								
RCC (12) *vs.* healthy controls (11)	75%	81.8%	81.8%	75%	83.3%	81.8%	83.3%	81.8%
RCC (12) *vs.* benigns/controls (20)	75%	60%	52.9%	80%	83.3%	85.0%	76.9%	89.5%

[a] PPV, positive predictive value. NPV, negative predictive value.
[b] The figures in parentheses represent the number in the group.

Table 4 *The results of the neural-network model based on peak presence/absence or intensity models trained on RCC versus normal healthy controls and benign samples combined*

The results are given for the training set and the initial blind set.

Sets	Presence/absence model				Intensity model			
	Sensitivity	Specificity	PPV[a]	NPV	Sensitivity	Specificity	PPV	NPV
Training								
RCC (48)[b] *vs.* benigns/controls (58)	100%	100%	100%	100%	100%	98.3%	98.0%	100%
Initial blind								
RCC (12) *vs.* healthy controls (11)	66.7%	63.6%	66.7%	63.6%	83.3%	81.8%	83.3%	81.8%
RCC (12) *vs.* benigns/controls (20)	66.7%	75%	61.5%	78.9%	83.3%	80.0%	71.4%	88.9%

[a] PPV, positive predictive value. NPV, negative predictive value.
[b] The figures in parentheses represent the number in the group.

several peaks in the lower part of the spectra occurred. Testing these samples against the original network produced a sensitivity of 84.2% (16 of 19) and specificity of only 60% (12 of 20) for these samples when 100% would have been expected given their utilization in the original training set. It is of note that although profiles were similar, the intensities of the peaks >10,000 Da were still less than previously obtained; this was not because of degradation, as the spiked internal calibrant cytochrome C was also affected. Initial experiments indicate that adjustment of the detector voltage from the reset values of the new detector may rectify this.

DISCUSSION

A key challenge in cancer medicine is to detect the disease as early as possible to allow the maximum chance of early therapeutic intervention. The use of multiple biomarkers has been suggested as possibly being the most promising way forward in terms of enhancing the positive predictive value of a test and minimizing false-positive and false-negative results (1). Therefore, technological approaches such as SELDI offer great potential and complement two-dimensional PAGE in terms of being able to provide precise molecular size profiling data on proteins or peptides smaller than many of those normally present on gels. Several studies using serum have now been published where the combination of SELDI profiling with computer-based modeling variably using genetic algorithms, cluster analysis, clustering, or boosted decision tree classification algorithms has produced extremely promising results. In breast, ovarian, and prostate cancers, sensitivities and specificities of 90–100% have been reported (31–35) together with a possible use in classifying men with marginally elevated prostate-specific antigen levels and of detecting prostate cancer at an early stage (31). In ovarian cancer, combined use of an iterative clustering algorithm with SELDI profiling of serum on H4 chips yielded a sensitivity of 100% and specificity of 95% based on initial training sets of 50 ovarian cancer patients and 50 unaffected women (healthy controls and benign ovarian conditions) with a subsequent blind set of 50 ovarian cancer patients, including 18 with stage I disease and 66 unaffected women (35). This exciting advance in being able to detect such early stage disease in a cancer, which most often is detected relatively late and at an advanced stage, has led to the proposal that this approach should be taken forward as a screening tool (35) both for high-risk and general populations. The majority of these SELDI-based studies thus far have examined the diagnostic utility of the approach, but, equally, other potential uses such as predicting or monitoring response to therapy are likely to be fruitful.

The establishment of suitably promising initial neural-network models in our study indicated the potential of this approach in being able to use urine samples to discriminate between renal cancer and control sample groups. However, the sensitivity and specificity values, and ultimately the associated predictive values obtained in the initial small test group analyzed blind, indicated that our procedure would be unsuitable as a screening tool without additional refinement of the model. However, potential peaks of interest for subsequent characterization and quantification using assays such as immunoassays were seen. Although poor, the sensitivity achieved was better than that found in preliminary studies for NMP-22 in renal cancer, for example, where only 40% is achieved (20). The only previous studies using SELDI to analyze urine samples examined renal function after administration of radiocontrast medium to rats and humans undergoing cardiac catheterization (38) and profiled samples from 30 patients with transitional cell carcinoma (29). The reproducibility of the mass accuracy was similar in these studies to that reported here, although the reproducibility of peak intensities was not reported. However, CVs of 15–20% have been seen using serum samples after normalization of peak intensities (34). Similar to our findings, significant interindividual heterogeneity in urinary profiles was also seen in the TCC study (29) using SAX2 anion exchange chips, and with similar numbers of peaks and multiple protein changes detected in the cancer group including five potential novel biomarkers. Using individual peaks, sensitivity ranged from 43% to 70% and specificity from 70% to 86%, but using the presence of particular clusters of peaks, sensitivity was significantly increased to 87% with a specificity of 66%, with the detection of low-grade tumors being superior to conventional cytologic approaches. Using antibodies, peaks at 3.3/3.4 kDa were identified as being human defensins α1 and α2. Relatively little is known about the identities of smaller molecular weight proteins in urine, as they are not normally present on gels because of their size and also the fact that an initial dialysis step with cutoff of 10 kDa is often used in urine processing. In addition to the urinary tumor-associated proteins described previously, a number of proteins with proapoptotic and antitumor activity have also been found in urine (39), which may represent some of the peaks seen here. These include

Table 5 *The results of the peak absence/presence and semiquantitative peak intensity neural-network models trained on RCC versus normal healthy controls for the late blind set based on samples analyzed 10 months after the models were trained and established*

Sets	Presence/absence model				Intensity model			
	Sensitivity	Specificity	PPV[a]	NPV	Sensitivity	Specificity	PPV	NPV
Late Blind								
RCC (36)[b] *vs.* healthy controls (31)	38.9%	77.4%	66.7%	52.2%	61.1%	54.8%	61.1%	54.8%
RCC (36) *vs.* benigns/controls (44)	38.9%	79.5%	60.9%	61.4%	61.1%	70.5%	62.9%	68.9%

[a] PPV, positive predictive value. NPV, negative predictive value.
[b] The figures in parentheses represent the number in the group.

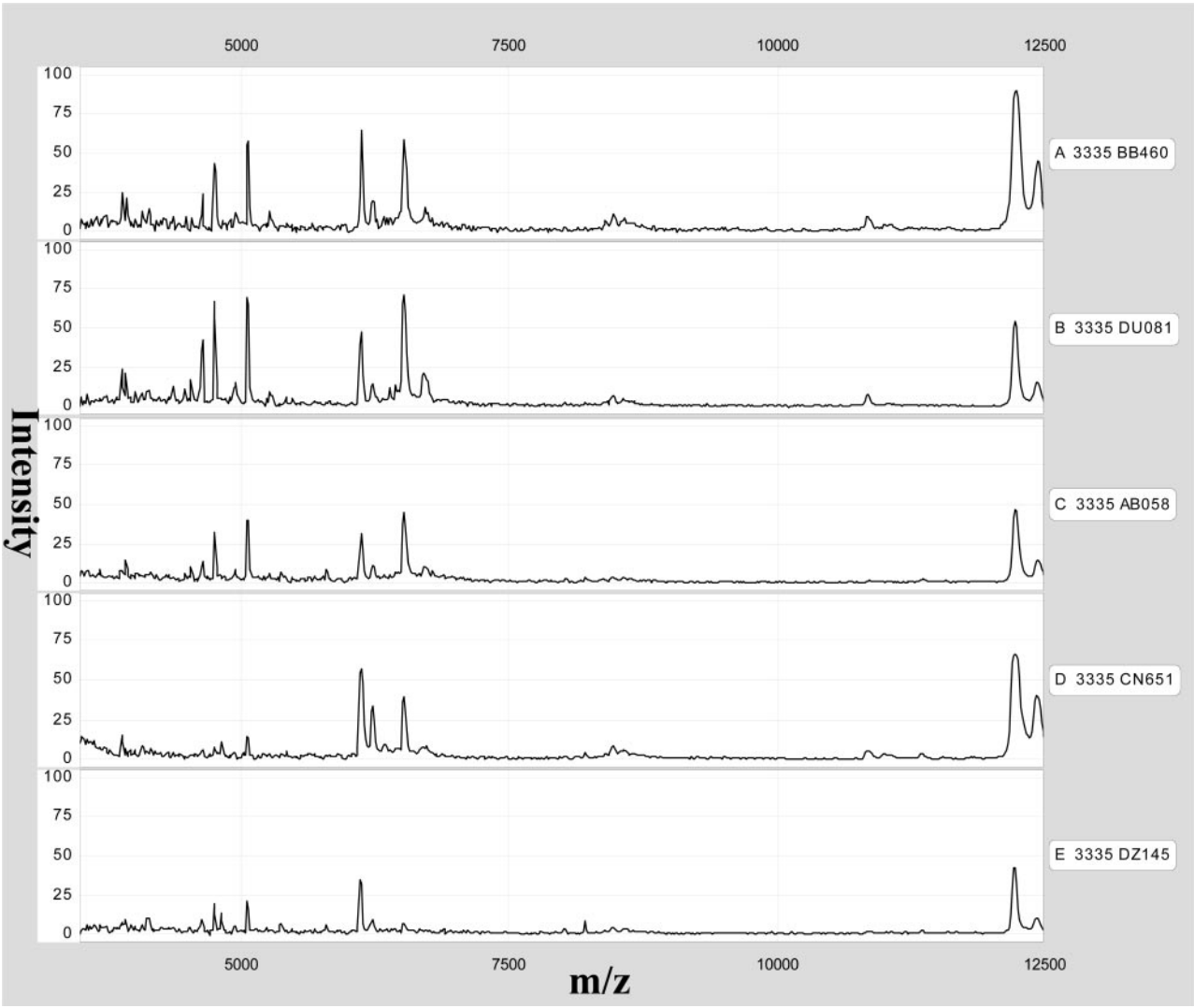

Fig. 4. Examples of spectra illustrating differences between different WCX2 chip batches for a single urine sample 3335. Spectrum A shows the initial result on the chip BB460 run as part of the initial training set using the settings shown in the text. The remaining spectra on chips DU081, AB058, CN651, and DZ145 all from different batches were generated ~1 year later using settings that had been adjusted to account for the decay in intensity because of instrument performance with laser intensity of 215 and detector voltage of 2200 V.

eosinophil-derived neurotoxin, antineoplastic urinary protein, angiostatin, inhibin, and activin A. A tentative identification of the presence of the peptide β defensin-1 in urine profiles in this study can be made based on comparison with previous non-SELDI-based urinary studies. Produced by many epithelial cell types including particularly the renal distal tubules and Loops of Henle in the kidney (40, 41), the presence of β defensin-1 in urine has been described previously (40, 42), with proforms and proteolytically processed forms including of 5068.1, 5068.7, 4749.2, and 4750.5 Da, very close to the masses of the peaks seen here at 5071.5 and 4753.3, and able to bind WCX2 chips (43). The reason for the decrease or relative absence in RCC urines may reflect increased proteolysis or alternatively decreased production, because complete loss of defensin expression has been shown in clear cell tumors (44).

However, the most marked problem in the present study is the poor results achieved with the subsequent larger blind set analyzed 10 months later, particularly the marked decline in sensitivity. There are several possible explanations ranging from sample processing and stability through machine performance to problems with the neural-network model. The neural-network model was used in combination with an equally promising peak detection algorithm that enabled the explicit identification of predictively relevant peaks, in contrast with the "black box" approaches often used in machine learning. As is typical with neural-network analyses, various aspects of this design were arrived at via a process of trial and error. Experimentation on fictional data sets was used to establish that the network could successfully classify cases given a few predictively relevant peaks against a background of noise, and experimentation within the training set determined the optimum number of inputs for each model type with this number determining the sizes of models used to predict the blind test sets. Peaks were ranked by their associated χ^2 statistics when cross-tabulated with the incidence of cancer, and the number of peaks used as input ranged from the best 10 to the best 70. The exact number of inputs was chosen after testing to find the smallest subset that would result in the minimum classification error. The choice of 5 hidden-layer neurons was somewhat arbitrary; the performance of networks with 10 neurons in the hidden layer was also examined but produced no improvement in predictive accuracy. More fundamentally, the use of a small hidden layer was intended as a defense against the dangers of over-fitting: with enough inputs and a large enough hidden layer, a network could be trained to successfully classify any data set. Given the modest sample size in comparison with the amount of data collected per subject, it seemed prudent to limit the space of possible models. Although a 5-hidden-neuron model will fail to cap-

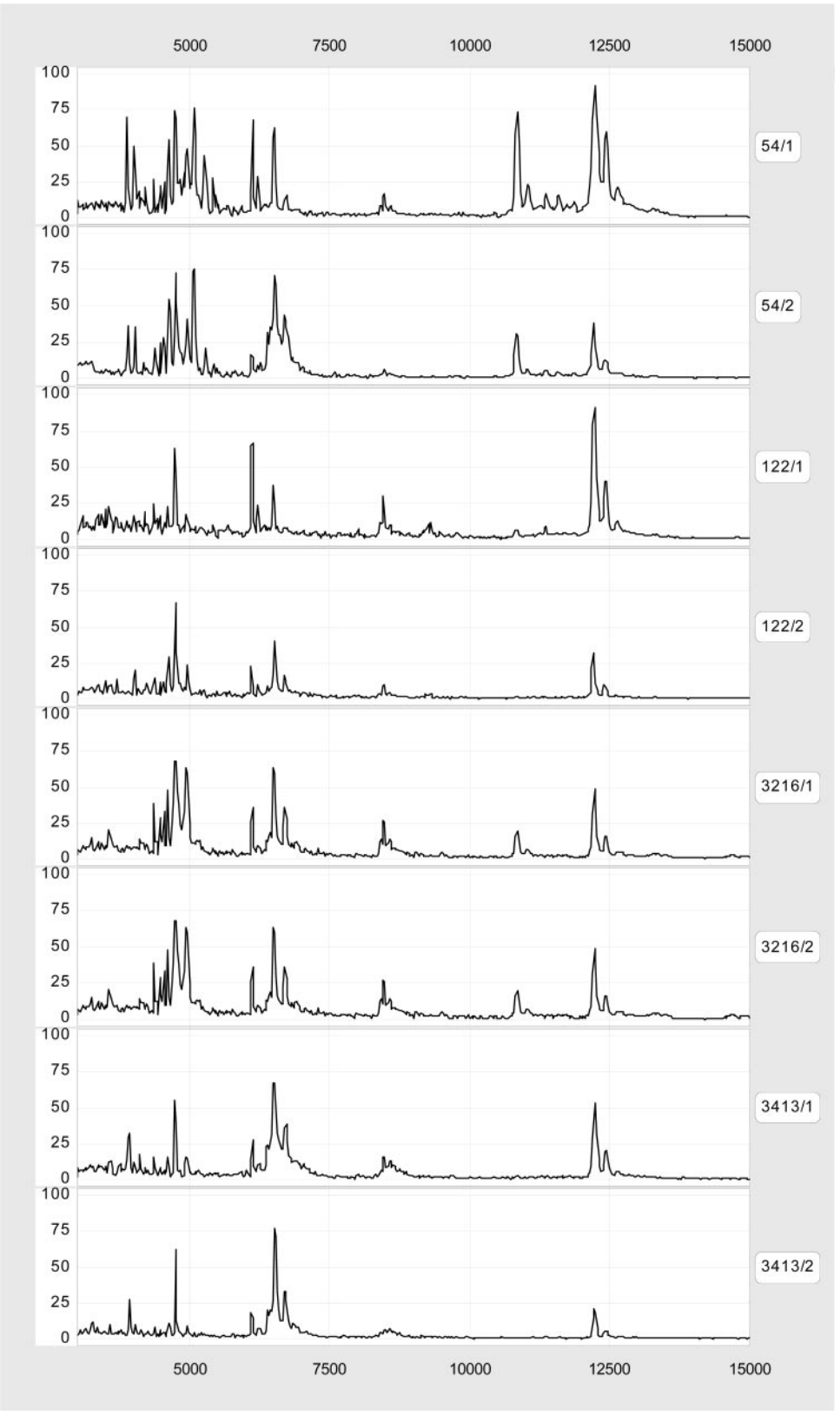

Fig. 5. Representative examples of SELDI WCX2 spectra from urine samples used in the neural-network training set (laser intensity 210) and subsequently rerun 18 months later after laser replacement (laser intensity 190).

ture highly complex interactions between input variables, it also prevents the detection of spurious interactions through over-fitting to the training set. (By way of comparison with traditional statistical techniques, the use of 5 neurons in the hidden layer is roughly equivalent to limiting a factor analysis or principal components analysis to the top 5 factors). An element of over-fitting may have occurred with the artificial neural network, which would account for the 15% drop in predictive accuracy with the initial blind test samples, but we feel that this is then unlikely to account for an additional drop 10 months later, although there may be a minor contribution.

An area for future development would be the comparative use of different modeling strategies and decisions regarding optimal samples numbers on which to establish models. Sample group size is critical, but the method of selection is not clear with traditional sample size calculations not necessarily applying to this kind of data where multiple biomarkers are being analyzed by machine learning approaches (45). By iteratively varying the numbers of subjects included in the test and training sets, suitable numbers should be found beyond which the power of the model does not substantially improve. Whatever the technique used to build predictive models, experimentation will always be required to fine-tune the approach. For example, using serum-based SELDI datasets, a sensitivity of 83% and specificity of 97% for prostate cancer *versus* noncancer was found using a simple decision tree classification algorithm (32) but improved to 95.8–100% on the same datasets if a boosted decision tree algorithm was used (34).

Precautions were taken to ensure that issues such as urine stability and processing would contribute minimally in introducing variability to the study. Stability of urine samples during freezing and processing may be more of an issue than for plasma or serum samples with the potential for increased protease content particularly in patients with cancer and the relative lack of endogenous protease inhibitors compared with serum. All of the samples were processed and stored identically and although sample stability was not generally found to be a problem, the use of protease inhibitor mixtures is a precaution that we use routinely and that does not appear to interfere with analysis by SELDI with the exception of introducing an aprotinin peak at 6504 Da. The amount used was titrated on the basis of ensuring minimal interference of this peak on the spectra while being within the recommended working range. The issue of differences in sample stability accounting for some of the problems seen here cannot be ruled out with storage times varying from weeks to 2 years. However, all of the samples were stored at −70 to −80°C, which has been shown to be generally less detrimental to several urinary proteins than −20°C (46), and, although samples in the control groups tended to have been stored for less time than the RCC groups, samples in both the training sets and late blind groups contained similar mixtures of samples in terms of storage times so that the possibility of differential sample storage artifacts was unlikely. However, samples were always diluted and applied to the chip with minimal delay, because changes in profiles were occasionally seen in samples that had been diluted and purposely left on ice for prolonged periods of time before analysis (4 h), but they appeared to be nonselective and may represent precipitation. There is clearly a need to standardize protein amount loaded, a practice that is not always followed in the published studies using SELDI and, indeed, may not always be necessary depending on the sample type if sufficient sample is applied to exceed the protein threshold. Therefore, differences in protein concentration could not account for the results seen and, although factors such as ionic strength of solutions are important in determining ion-exchange binding, urinary composition in terms of salt and urea concentrations were not found to affect profiles on WCX2 chips within the ranges examined. The demonstrated necessity to standardize protein concentrations precluded normalizing for hydration state using creatinine concentrations, as would normally happen for a quantitative assay of any analyte. A mathematical adjustment of peak intensities taking into account protein and creatinine ratios of samples before analysis was investigated, but this did not improve the power of the models used. This is not surprising, because the correction for any sample could only be applied uniformly to peaks detected, and these are partly governed by amount of sample loaded that, if normalized for creatinine rather than protein, may change between samples.

Machine performance in terms of laser intensity changes clearly alters with time and, it is likely that this is at least partly implicated in the failure of the model when used over a long period of time. Although the primary failure in the blind set was most marked in sensitivity, adjustment of the replacement laser and reanalysis of 39 training set samples clearly boosted the results toward better sensitivity rather than specificity. However, the failure to achieve 100% sensitivity and specificity clearly indicates that other factors or adjustments need to be made. Laser performance will deteriorate with time, and quite clearly measures need to be taken to monitor this regularly and adjust settings accordingly. The use of a standard chip, which can be used for calibration and to monitor laser performance in terms of the peak intensities, may be ideal, because this would not rely on chip chemistry, being essentially dried on to the chip surface. We are currently examining various ways of achieving this monitoring and being able to detect subtle changes over time, in addition to those readily apparent when using this approach in our case when a study is recommenced after a long break. An automatic laser adjustment feature is apparently now included in the latest version of the Ciphergen software, but whether this can achieve the reproducibility of performance needed remains to be determined. In addition, changes in the detector performance with time may be important. The issue of quantification by mass spectrometry is technically very demanding and when used as in this and in similar studies, can at best be regarded as semiquantitative, with a variety of factors influencing protein ionization including the presence of other proteins in the mixture analyzed exerting suppressive effects. Normalization of peaks against total ion current may be one software option to improve this aspect, but again for samples such as urine with relatively few peaks and exhibiting a relatively high degree of heterogeneity, this may not be as valid as for more complex samples and certainly could not be used to improve long-term comparisons.

The variability found with some chips, which may be batch-related on the basis of the results shown here but also may represent variability of some chip spots within batches, although not found within our initial reproducibility studies, is also likely to have contributed to the results. This requires additional evaluation, and the use of duplicate analyses for example may be explored but our initial work indicated that this would not necessarily overcome the issue. This has not been highlighted before, although in one of the prostate cancer trials quality control chips consisting of pooled normal and cancer serum samples were developed to allow monitoring of between-day/chip variability (34). We are also exploring similar practices to assess chip chemistry combining this with the standard chips to assess laser performance as described. To our knowledge, no studies examining the long-term use of computational profiling in this scenario have been published, although one study using serum on immobilized metal affinity chromatography chips has reported that randomly selected samples rerun months or even a year after the study were correctly classified (32). It is not clear how many samples were examined in this way or whether the same batch of chips was used.

The results initially obtained here for renal cancer are equal to or superior to the urinary protein assays described currently for bladder cancer (17, 18), and even the later blind results are similar to those obtained with BTA and BTA Stat. However, such assays have a role in monitoring recurrence in bladder cancer where local recurrence is common, whereas for renal cancer, recurrence at other sites is more likely and, therefore, serum-based markers may be more useful. However, the more important issues raised by this study are the implications for the long-term robustness of the type of approach used here. With the recognition that no single biomarker fulfils all of the desirable roles of markers in terms of utility in diagnosis, prognosis, or monitoring, approaches using multiple parameters such as SELDI and other similar array-based approaches, whether at the cDNA or

protein level, may ultimately prove to be the most versatile. It could be that some of the problems experienced in this study may have been compounded by the use of urine, which contains far fewer peaks and is potentially more labile. Several extremely promising studies have now been published using SELDI in combination with computer algorithms and serum samples, although follow-up studies regarding long-term utility have not yet been published. If they prove to be more robust in the long-term, then this may be an indication that serum samples may be more appropriate. The exciting challenge then will be their translation to clinical practice with prospective samples in multiple geographic sites. There is concern about the designs of trials to formally take promising biomarkers forward and the importance of maintaining standards (47–49). Several initiatives such as the Tumor Marker Utility Grading System (TMUGS), the Early Detection Research Network (EDRN), and the Program for Assessment of Clinical Cancer Tests (PACCT) address issues ranging from coordination of research activities through to experimental validation and to guidelines for clinical evaluation (45, 48, 50, 51). The specific technical factors highlighted in this study are likely to be important considerations in the pursuit of successful translation of SELDI-based approaches in this area.

ACKNOWLEDGMENTS

We thank the clinical urology teams for support.

REFERENCES

1. Srinivas, P. R., Kramer, B. S., and Srivastava, S. Trends in biomarker research for cancer detection. Lancet Oncol., *2:* 698–704, 2001.
2. Vogelzang, N. J., and Stadler, W. M. Kidney cancer. Lancet, *352:* 1691–1696, 1998.
3. Pantuck, A. J., Zisman, A., and Belldegrun, A. S. The changing natural history of renal cell carcinoma. J. Urol., *166:* 1611–1623, 2001.
4. Glaspy, J. A. Therapeutic options in the management of renal cell carcinoma. Semin. Oncol., *29:* 41–46, 2002.
5. Zhou, M., and Rubin, M. A. Molecular markers for renal cell carcinoma: impact on diagnosis and treatment. Semin. Urol. Oncol., *19:* 80–87, 2001.
6. Anderson, N. G., Anderson, N. L., and Tollaksen, S. L. Proteins of human urine. I. Concentration and analysis by two-dimensional electrophoresis. Clin. Chem., *25:* 1199–1210, 1979.
7. Edwards, J. J., Tollaksen, S. L., and Anderson, N. G. Proteins of human urine. III. Identification and two-dimensional electrophoretic map positions of some major urinary proteins. Clin. Chem., *28:* 941–948, 1982.
8. Marshall, T., Williams, K. M., and Vesterberg, O. Unconcentrated human urinary proteins analysed by high resolution two-dimensional electrophoresis with narrow pH gradients: preliminary findings after occupational exposure to cadmium. Electrophoresis, *6:* 47–52, 1985.
9. Büeler, M. R., Wiederkehr, F., and Vonderschmitt, D. J. Electrophoretic, chromatographic and immunological studies of human urinary proteins. Electrophoresis, *16:* 124–134, 1995.
10. Grover, P. K., and Resnick, M. I. Two-dimensional analysis of proteins in unprocessed human urine using double stain. J. Urol., *150:* 1069–1072, 1993.
11. Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen, K., Wahl, R. C., Yu, W., Luethy, R., and Patterson, S. D. Towards defining the urinary proteome using liquid chromatography-tandem mass spectrometry. I. Profiling an unfractionated tryptic digest. Proteomics, *1:* 93–107, 2001.
12. Pang, J. X., Ginanni, N., Dongre, A. R., Hefta, S. A., and Opiteck, G. J. Biomarker discovery in urine by proteomics. J. Proteome. Res., *1:* 161–169, 2002.
13. Edwards, J. J., Anderson, N. G., Tollaksen, S. L., von Eschenbach, A. C., and Guevara, J., Jr. Proteins of human urine. II. Identification by two-dimensional electrophoresis of a new candidate marker for prostatic cancer. Clin. Chem., *28:* 160–163, 1982.
14. Ostergaard, M., Wolf, H., Orntoft, T. F., and Celis, J. E. Psoriasin (S100A7): a putative urinary marker for the follow-up of patients with bladder squamous cell carcinomas. Electrophoresis, *20:* 349–354, 1999.
15. Rasmussen, H. H., Orntoft, T. F., Wolf, H., and Celis, J. E. Towards a comprehensive database of proteins from the urine of patients with bladder cancer. J. Urol., *155:* 2113–2119, 1996.
16. Stenman, U. H. Tumor-associated trypsin inhibitor. Clin. Chem., *48:* 1206–1209, 2002.
17. Konety, B. R., and Getzenberg, R. H. Urine based markers of urological malignancy. J. Urol., *165:* 600–611, 2001.
18. Ross, J. S., and Cohen, M. B. Biomarkers for the detection of bladder cancer. Adv. Anat. Pathol., *8:* 37–45, 2001.
19. Huang, S., Rhee, E., Patel, H., Park, E., and Kaswick, J. Urinary NMP22 and renal cell carcinoma. Urology, *55:* 227–230, 2000.
20. Ozer, G., Altinel, M., Kocak, B., Yazicioglu, A., and Gonenc, F. Value of urinary NMP-22 in patients with renal cell carcinoma. Urology, *60:* 593–597, 2002.
21. Rodriguez-Cuartero, A., Barcelona-Martin, F., and Perez-Blanco, F. J. Urinary β-glucuronidase in renal cell carcinoma. Clin. Nephrology, *53:* 288–290, 2000.
22. Merchant, M., and Weinberger, S. R. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis, *21:* 1164–1177, 2000.
23. von Eggeling, F., Junker, K., Fiedle, W., Wollscheid, V., Durst, M., Claussen, U., and Ernst, G. Mass spectrometry meets chip technology: a new proteomic tool in cancer research? Electrophoresis, *22:* 2898–2902, 2001.
24. Paweletz, C. P., Gillespie, J. W., Ornstein, D. K., Simone, N. L., Brown, M. R., Cole, K. A., Wang, Q-H., Huang, J., Hu, N., Yip, T-T., Rich, W. E., Kohn, E. C., Linehan, W. M., Weber, T., Taylor, P., Emmert-Buck, M. R., Liotta, L. A., and Petricoin, E. F. Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip. Drug Dev. Res., *49:* 34–42, 2000.
25. Wellmann, A., Wollscheid, V., Lu, H., Ma, Z. L., Albers, P., Schutze, K., Rohde, V., Behrens, P., Dreschers, S., Ko, Y., and Wernert, N. Analysis of microdissected prostate tissue with ProteinChip arrays - a way to new insights into carcinogenesis and to diagnostic tools. Int. J. Mol. Med., *9:* 341–347, 2002.
26. Cazares, L. H., Adam, B. L., Ward, M. D., Nasim, S., Schellhammer, P. F., Semmes, O. J., and Wright, G. L., Jr. Normal, benign, preneoplastic, and malignant prostate cells have distinct protein expression profiles resolved by surface enhanced laser desorption/ionization mass spectrometry. Clin. Cancer Res., *8:* 2541–2552, 2002.
27. Wright Jr., G. L., Cazares, L. H., Leung, S-M., Nasim, S., Adam, B. L., Yip, T-T., Schellhammer, P. F., Gong, L., and Vlahou, A. Proteinchip surface enhanced laser desorption/ionization (SELDI) mass spectrometry: a novel protein biochip technology for detection of prostate cancer biomarkers in complex protein mixtures. Prostate Cancer Prostatic Dis., *2:* 264–276, 1999.
28. Xiao, Z., Adam, B. L., Cazares, L. H., Clements, M. A., Davis, J. W., Schellhammer, P. F., Dalmasso, E. A., and Wright, G. L., Jr. Quantitation of serum prostate-specific membrane antigen by a novel protein biochip immunoassay discriminates benign from malignant prostate disease. Cancer Res., *61:* 6029–6033, 2001.
29. Vlahou, A., Schellhammer, P. F., Mendrinos, S., Patel, K., Kondylis, F. I., Gong, L., Nasim, S., and Wright Jr., G. L. 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.
30. Rosty, C., Christa, L., Kuzdzal, S., Baldwin, W. M., Zahurak, M. L., Carnot, F., Chan, D. W., Canto, M., Lillemoe, K. D., Cameron, J. L., Yeo, C. J., Hruban, R. H., and Goggins, M. Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein I as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. Cancer Res., *62:* 1868–1875, 2002.
31. Petricoin, E. F., III, Ornstein, D. K., Paweletz, C. P., Ardekani, A., Hackett, P. S., Hitt, B. A., Velassco, A., Trucco, C., Wiegand, L., Wood, K., Simone, C. B., Levine, P. J., Linehan, W. M., Emmert-Buck, M. R., Steinberg, S. M., Kohn, E. C., and Liotta, L. A. Serum proteomic patterns for detection of prostate cancer. J. Natl. Cancer Inst., *94:* 1576–1578, 2002.
32. Adam, B. L., Qu, Y., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O. J., Schellhammer, P. F., Yasui, Y., Feng, Z., and Wright Jr., G. L. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res., *62:* 3609–3614, 2002.
33. Li, J., Zhang, Z., Rosenzweig, J., Wang, Y. Y., and Chan, D. W. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin. Chem., *48:* 1296–1304, 2002.
34. Qu, Y., Adam, B-L., Yasui, Y., Ward, M. D., Cazares, L. H., Schellhammer, P. F., Feng, Z., Semmes, O. J., and Wright, G. L. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. Clin. Chem., *48:* 1835–1843, 2002.
35. Petricoin III, E. F., Ardekani, 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.
36. Townsend, J. C., Sadler, W. A., and Shanks, G. M. The effect of storage pH on the precipitation of proteins in deep frozen urine samples. Ann. Clin. Biochem., *24(Pt 1):* 111–112, 1987.
37. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning internal representations by error propagation. *In:* D. E. Rumelhart and J. L. McClelland (eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, pp. 318–362. Cambridge, MA: MIT Press, 1986.
38. Hampel, D. J., Sansome, C., Sha, M., Brodsky, S., Lawson, W. E., and Goligorsky, M. S. Toward proteomics in uroscopy: urinary protein profiles after radiocontrast medium administration. J. Am. Soc. Nephrol., *12:* 1026–1035, 2001.
39. Pati, S., Lee, Y., and Samaniego, F. Urinary proteins with pro-apoptotic and antitumor activity. Apoptosis, *5:* 21–28, 2000.
40. Valore, E. V., Park, C. H., Quayle, A. J., Wiles, K. R., McCray, P. B., and Ganz, T. Human b-defensin-1: an antimicrobial peptide of urogenital tissues. J. Clin. Investig., *101:* 1633–1642, 1998.
41. Zucht, H. D., Grabowsky, J., Schrader, M., Liepke, C., Jurgens, M., Schulz-Knappe, P., and Forssmann, W. G. Human β-defensin-1: a urinary peptide present in variant molecular forms and its putative functional implication. Eur. J. Med. Res., *3:* 315–323, 1998.

42. Hiratsuka, T., Nakazato, M., Ihi, T., Minematsu, T., Chino, N., Nakanishi, T., Shimizu, A., Kangawa, K., and Matsukura, S. Structural analysis of human b-defensin-1 and its significance in urinary tract infection. Nephron, *85:* 34–40, 1999.
43. Diamond, D. L., Kimball, J. R., Krisanaprakornkit, S., Ganz, T., and Dale, B. A. Detection of β-defensins secreted by human oral epithelial cells. J. Immunol. Methods, *256:* 65–76, 2001.
44. Young, A. N., De Oliveira Salles, P. G., Lim, S. D., Cohen, C., Petros, J. A., Marshall, F. F., Neish, A. S., and Amin, M. B. β defensin-1, parvalbumin and vimentin: a panel of diagnostic immunohistochemical markers for renal tumors derived from gene expression profiling studies using cDNA microarrays. Am. J. Surg. Pathol., *27:* 199–205, 2003.
45. Sullivan, P. M., Etzioni, R., Feng, Z., Potter, J. D., Thompson, M. L., Thornquist, M., Winget, M., and Yasui, Y. Phases of biomarker development for early detection of cancer. J. Natl. Cancer Inst., *93:* 1054–1061, 2001.
46. Schultz, C. J., Dalton, R. N., Turner, C., Neil, H. A. W., and Dunger, D. B. Freezing method affects the concentration and variability of urine proteins and the interpretation of data on microalbuminuria. Diabetic Med., *17:* 7–14, 2000.
47. Sargent, D., and Allegra, C. Issues in clinical trial design for tumor marker studies. Semin. Oncol., *29:* 222–230, 2002.
48. Hammond, M. E., and Taube, S. E. Issues and barriers to development of clinically useful tumor markers: a development pathway proposal. Semin. Oncol., *29:* 213–221, 2002.
49. Biomarkers Ad-hoc Group of the United Kingdom Coordinating Committee on Cancer Research. Biological markers: maintaining standards. Br. J. Cancer, *82:* 1627–1628, 2000.
50. Hayes, D. F., Bast, R. C., Desch, C. E., Fritsche, H., Jr., Kemeny, N. E., Jessup, J. M., Locker, G. Y., Macdonald, J. S., Mennel, R. G., Norton, L., Ravdin, P., Taube, S., and Winn, R. J. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. J. Natl. Cancer Inst., *88:* 1456–1466, 1996.
51. Verma, M., Wright, G. L., Jr., Hanash, S. M., Gopal-Srivastava, R., and Srivastava, S. Proteomic approaches within the NCI early detection research network for the discovery and identification of cancer biomarkers. Ann. N. Y. Acad. Sci., *945:* 103–115, 2001.