

## Serum Proteomic Fingerprinting Discriminates Between Clinical Stages and Predicts Disease Progression in Melanoma Patients

Shahid Mian, Selma Ugurel, Erika Parkinson, Iris Schlenzka, Ian Dryden, Lee Lancashire, Graham Ball, Colin Creaser, Robert Rees, and Dirk Schadendorf

From the Interdisciplinary Biomedical Research Centre, School of Science, Nottingham Trent University, Clifton Lane, Clifton; School of Mathematical Sciences, University of Nottingham, University Park, Nottingham; Department of Animal Physiology, School of Biosciences, University of Nottingham, Sutton Bonington, Loughborough, United Kingdom; Skin Cancer Unit, German Cancer Research Center, Heidelberg; Department of Dermatology, University Hospital, Mannheim, Germany.

Submitted March 27, 2004; accepted November 19, 2004.

Supported by the European Union grant QLGI-CT-2002-00668 (OISTER) and the John and Lucille van Geest Foundation.

All authors have contributed equally to this manuscript.

Authors' disclosures of potential conflicts of interest are found at the end of this article.

Address reprint requests to Professor Robert Rees, Interdisciplinary Biomedical Research Centre, School of Science, Nottingham Trent University, Clifton Lane, Clifton, Nottingham NG11 8NS, United Kingdom; e-mail: robert.rees@ntu.ac.uk.

© 2005 by American Society of Clinical Oncology

0732-183X/05/2322-5088/\$20.00

DOI: 10.1200/JCO.2005.03.164

### ABSTRACT

#### Purpose

Currently known serum biomarkers do not predict clinical outcome in melanoma. S100- $\beta$  is widely established as a reliable prognostic indicator in patients with advanced metastatic disease but is of limited predictive value in tumor-free patients. This study was aimed to determine whether molecular profiling of the serum proteome could discriminate between early- and late-stage melanoma and predict disease progression.

#### Patients and Methods

Two hundred five serum samples from 101 early-stage (American Joint Committee on Cancer [AJCC] stage I) and 104 advanced stage (AJCC stage IV) melanoma patients were analyzed by matrix-assisted laser desorption/ionisation (MALDI) time-of-flight (ToF; MALDI-ToF) mass spectrometry utilizing protein chip technology and artificial neural networks (ANN). Serum samples from 55 additional patients after complete dissection of regional lymph node metastases (AJCC stage III), with 28 of 55 patients relapsing within the first year of follow-up, were analyzed in an attempt to predict disease recurrence. Serum S100- $\beta$  was measured using a sandwich immunoluminometric assay.

#### Results

Analysis of 205 stage I/IV serum samples, utilizing a training set of 94 of 205 and a test set of 15 of 205 samples for 32 different ANN models, revealed correct stage assignment in 84 (88%) of 96 of a blind set of 96 of 205 serum samples. Forty-four (80%) of 55 stage III serum samples could be correctly assigned as progressors or nonprogressors using random sample cross-validation statistical methodologies. Twenty-three (82%) of 28 stage III progressors were correctly identified by MALDI-ToF combined with ANN, whereas only six (21%) of 28 could be detected by S100- $\beta$ .

#### Conclusion

Validation of these findings may enable proteomic profiling to become a valuable tool for identifying high-risk melanoma patients eligible for adjuvant therapeutic interventions.

*J Clin Oncol* 23:5088-5093. © 2005 by American Society of Clinical Oncology

### INTRODUCTION

Blood contains a plethora of undefined biomarkers that may reflect the state of the individual organism.<sup>1,2</sup> Recently, proteomic and bioinformatic approaches were shown to be able to dissect the serum proteome and identify signature biomarker patterns indicative of cancers of different origin (eg, ovary,

breast, and prostate).<sup>3,4</sup> The application of this approach, if validated for its sensitivity and robustness, may influence diagnostic and therapeutic decisions, as shown by the successful prediction of the chemoresponsiveness of breast cancer cell lines.<sup>5</sup> As no serum biomarker is currently known to reliably predict the prognosis of melanoma patients, the present study was designed to test

whether signature patterns of proteomic profiles obtained from serum samples could be indicative for different stages of the disease and predict its progression.

## PATIENTS AND METHODS

Serum samples were selected from a frozen collection of sera from patients with histologically confirmed melanoma. All serum samples were obtained and processed following a standardized protocol: Blood was drawn from patients' cubital veins into gel-coated serum tubes (Sarstedt, Nuembrecht, Germany) and allowed to clot at room temperature for at least 30 minutes, but for no longer than 60 minutes. Thereafter, the tubes were centrifugated at 2,500 g for 10 minutes. The serum phase was harvested and subsequently frozen without any additives in 1-mL aliquots at  $-20^{\circ}\text{C}$  and not thawed until immediately before analysis. The collection of sera and clinical data was performed after receipt of patients' informed consent with institutional review board approval. Serum samples from stage I patients were selected as follows: (1) the blood samples must have been obtained in a time frame of 2 to 6 weeks after surgical resection of the primary tumor, and (2) patients must have been confirmed to have stage I disease according to the guidelines of the American Joint Committee on Cancer<sup>6</sup> (in brief: primary melanoma with a tumor thickness up to 2.0 mm without ulceration or a tumor thickness up to 1.0 mm with ulceration or Clark's level IV or V, respectively; no evidence of metastatic disease). Serum samples from stage III patients were selected according to the following criteria: (1) confirmation of macroscopic stage III disease<sup>6</sup> (in brief: macrometastases of the regional lymph nodes and/or satellite or in-transit metastases; no evidence of distant metastases); (2) blood withdrawal within a time frame of 2 to 6 weeks after complete surgical dissection of the affected lymph node basin; (3) no systemic treatment in stage III prior to blood withdrawal; and (4) performance of follow-up examinations, including physical examination, chest x-ray/computed tomography (CT), ultrasound or CT of the abdomen and regional lymph nodes, as well as blood chemistry at regular intervals of 3 months for at least 1 year. Serum samples from stage IV patients were selected as follows: (1) confirmation of stage IV disease<sup>6</sup> (in brief distant metastases); and (2) no systemic treatment in stage IV prior to blood withdrawal.

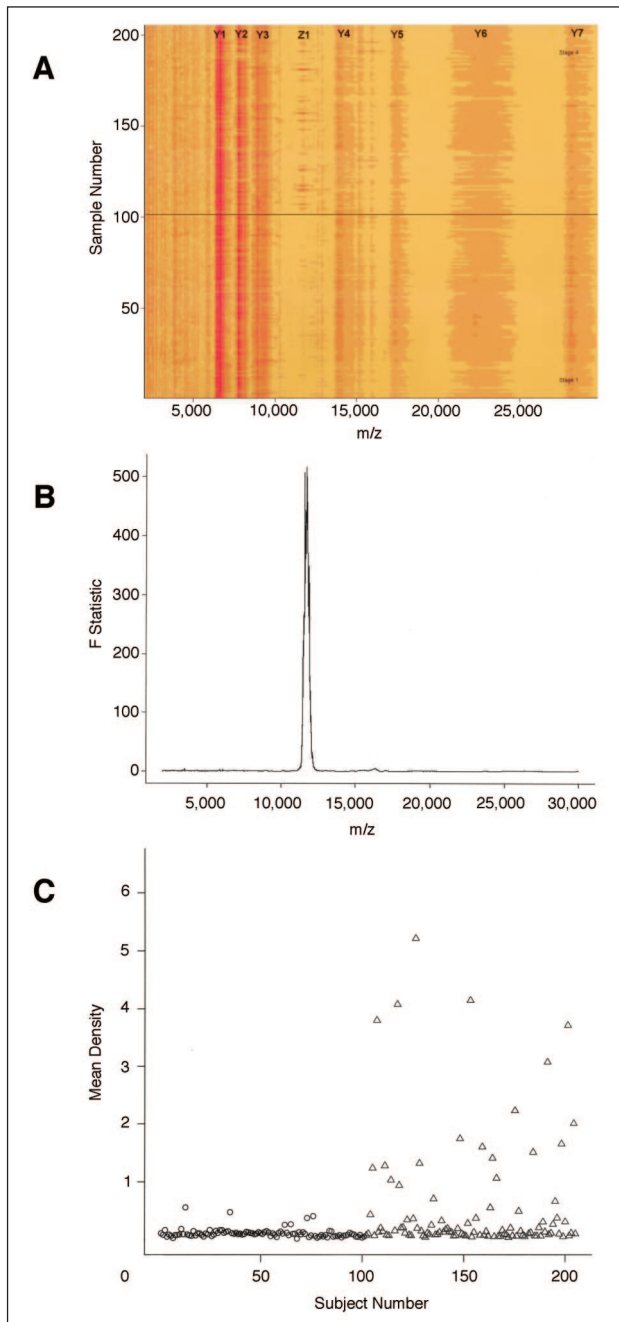
Matrix-assisted laser desorption/ionisation (MALDI) time-of-flight (ToF; MALDI-ToF) mass spectrometry analysis was conducted using a phosphate-buffered saline II mass analyser (Ciphergen Biosystems, Fremont, CA). Two  $\mu\text{L}$  of undiluted serum was subjected onto a H4 protein chip (Ciphergen Biosystems) and allowed to bind at room temperature for 15 minutes in a humidified chamber. Serum was removed, the chip surface was washed five times, and the surface was allowed to air dry. Thereafter, 0.8  $\mu\text{L}$  of a saturated solution of sinapinic acid was added to each spot. Mass analysis conditions were conducted as follows: laser intensity setting of 270, sensitivity of 6, with 65 transient collections per spot using automated data collection. Data acquisitions were made from 0 to 30 kDa, and mass accuracy was determined to be approximately 0.2% of actual mass values using external calibration with bovine superoxide dismutase single- and double-charged peaks. To check mass accuracy, serum samples were run as a single batch, with calibrants being placed on approximately one chip in every eight. Spectra were background subtracted and then exported as csv files into Microsoft Excel

(Microsoft, Redmond, WA). Density plots using log of the data were performed and plotted against sample numbers using the statistical program "R" (open source available at <http://www.T-project.org>). The obtained data were used to train and test supervised learning algorithms based on artificial neural networks (ANN; Neuroshell 2; Ward Systems, Frederick, MD). The ANN architecture is a three-layer multilayer perceptron with a sigmoidal transfer function on the hidden and output layers, and a linear function on the input layer. The algorithm used is a feed-forward/back-propagation network, more commonly termed a back-propagation network. Serum concentration of S100- $\beta$  was measured using a sandwich immunoluminometric assay (LIA-mat Sangtec 100; Sangtec Medical, Bromma, Sweden).

## RESULTS

An initial exploratory analysis of the data was carried out (Fig 1A) to examine reproducibility of the mass spectra of all 205 samples. Data from mass values of 2,000 to 30,000 Da were plotted on a log scale. The y-axis denotes the patient sample spectra obtained from stage I (numbers 1 to 101) and stage IV patients (numbers 102 to 205; note that stage I and stage IV patients are separated by a black line), whereas the x-axis displays the mass-charge ( $m/z$ ) value. Higher scan readings are denoted by dark pixel coloration as compared with lower scan readings, which are represented by light pixel coloration. Figure 1A suggests that there is no overall systematic difference between the stage I or stage IV populations (Fig 1A lanes Y1 to Y7). There is, however, one notable feature of the data set in which the presence of a signal with an average mass value of 11,700 Da produces higher intensity readings within a greater proportion of stage IV than stage I melanoma samples (Fig 1A lane Z1). We then evaluated the F statistic for two independent populations for each  $m/z$  value between 2,000 and 30,000 Da (Fig 1B). The most obvious feature is the huge value of the F statistic, around 11,700, as expected from the density map shown in Figure 1A. The peak with high F values is very wide. A further plot of the average value of the data over this range of  $m/z$  for each observation is shown in Figure 1C. The data indicate that approximately 25% of stage IV melanoma samples (triangles) have very high scan readings over this range as compared with the stage I sample spectra (circles).

The profiles of 205 serum samples, corresponding to 101 stage I and 104 stage IV patients, were imported into Neuroshell 2 and randomized, and a subset of 50 ( $n = 25$  stage I;  $n = 25$  stage IV) was selected as an independent validation data set. For the remaining 155 samples, 94 were utilized to train the ANN through an iterative learning process, with 15 samples used to test model performance (training was stopped when model performance failed to improve for 100,000 events), with the remaining 46 samples providing an additional blind data set ( $n = 96$  total



**Fig 1.** (A) Density plot of relative intensity data derived from mass spectral peak analysis versus patient sample number for stage I and stage IV patients. Log values of density for each mass-charge ratio ( $m/z$ ) from 2,000 to 30,000 Da (x-axis) were plotted against patient number (y-axis). Patients 1 to 101 represent stage I, and 102 to 205 represent stage IV. Higher intensity values are in red. Y1 to Y7 represents bands consistent between both populations. Z1 indicates a mass with higher intensity predominantly associated with stage IV patients. (B) F test for testing equality of two independent population variances; evaluation of the F score for testing between two independent population variances for all  $m/z$  values between 2,000 and 30,000 Da using the statistical program "R." Mass values centered at approximately 11,700 Da have been identified as highly significant. (C) Mean intensity readings for  $m/z$  values: 11,583 to 11,788 Da; a plot of mean intensity values in the range of 11,583 to 11,788 Da for each sample. Approximately 25% of stage IV samples ( $\Delta$ ) have higher readings compared with the stage I samples ( $\circ$ ). The y-axis denotes mean intensity values, with the x-axis corresponding to patient sample. Patients 1 to 101 and 102 to 205 represent stage I and stage IV patients, respectively.

blind samples). Receiver operating characteristic curves, sensitivity, specificity, positive predictive value (the ability to predict true-positives from false-positives), and negative predictive value (the ability to identify true-negatives from false-negatives) for both independent datasets ( $n = 96$  samples) are presented for the first eight ANN architectures (Table 1). An alternative method of prediction is to use Fisher's linear discriminant rule on the first few principal component scores from the data. Using 100 training sets of 50 randomly chosen observations from each group, averages of 71%, 78%, 80%, 81%, 83%, 84%, and 86% correct classification were obtained using the first 2, 3, 5, 10, 15, 20, and 30 principal component scores, respectively. So the ANN and discriminant analysis give broadly similar classification rates.

Next, 33 of 55 serum samples from stage III melanoma patients were chosen at random to train the ANN, 11 to test model performance during training with the remaining 11 samples chosen to test the predictive capability of the system for blind data via a random sample cross validation approach.<sup>5</sup> On error convergence the system was utilized to predict the staging class of the remaining 11 blind samples. This procedure was repeated a total of 51 times, thus enabling every sample to appear in training and blind sample sets, and allowed an average class assignment value for a particular serum sample to be obtained. Class assignment was derived using a Student's  $t$  test at the 5% significance level. Patients progressing within 1 year of clinical follow-up were given discriminatory numerical values of 1, and those not progressing to stage IV were assigned values of 2. Data analysis indicated that 44 (80%) of 55 correct class assignments were obtained, with 39 of 44 occurring at  $P < .05$  (Fig 2). Assessment of S100- $\beta$  revealed elevated concentrations in six (21%) of 28 progressors and four (15%) of 27 non-progressors (Fig 2).

## DISCUSSION

The identification of melanoma patients who are at risk of disease progression is an essential task of appropriate clinical management. Serum markers of predictive relevance are few.<sup>7-9</sup> The widely used S100- $\beta$  correlates with tumor burden and, therefore, is of limited predictive value in tumor-free patients.<sup>10-12</sup> Our results show that nonlinear ANN bioinformatic algorithms, in conjunction with protein profiling technologies, can discriminate between serum protein expression patterns from either stage I or stage IV melanoma patients for accurate disease staging. Data parameterization, in terms of hierarchical ranking, led to the identification of key ions that could be important predictive biomarkers. At least, these technologies offer the potential to identify patients at enhanced risk of disease progression even in patients with no detectable macroscopic disease. The low predictive sensitivity of serum S100- $\beta$  (six of 28;

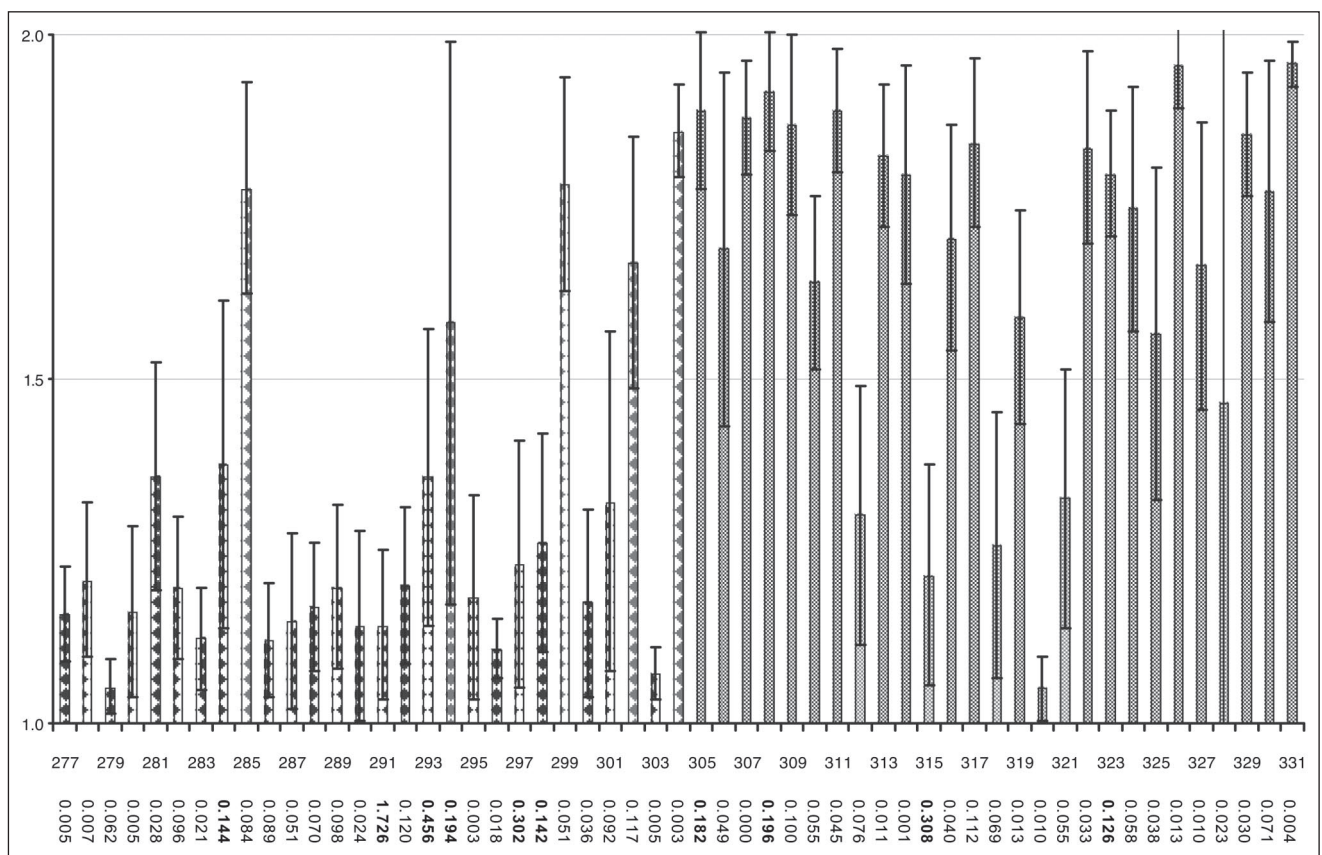
**Table 1.** Classification Performance for the First Eight Artificial Neural Networks Architectures in Predicting Class Assignment for 96 Blind Stage I and Stage IV Melanoma Samples

Model No.	Accuracy	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Area Under the Curve
1	88	84	92	91.30434783	85.18518519	0.9512
2	86	84	88	87.5	84.61538462	0.9368
3	86	84	88	87.5	84.61538462	0.9368
4	86	80	92	90.90909091	82.14285714	0.948
5	88	88	88	88	88	0.9048
6	80	80	80	80	80	0.884
7	78	76	80	79.16666667	76.92307692	0.8664
8	80	88	72	75.86206897	85.71428571	0.8544

NOTE. Stage IV samples are treated as positive, and stage I as negative. Sensitivity indicates the percentage of stage IV serum samples that are correctly classified. Specificity represents percentage of stage I samples correctly classified. Positive predictive values represent percentage of true-positives distinguished from false-positives. Negative predictive values represent percentage of true-negatives from false-negatives.

21%) compared with MALDI-ToF combined with ANN (23 of 28; 82%) to identify stage III progressors implies that other proteins than S100-β are major components of the detected signature profiles indicative of disease progression. While conventional clinical assessment was performed on these stage III patients for 1 year, it would have been useful

to collect and analyze serum samples from these patients in order to ascertain whether discrimination between stage III patients was consistent and reproducible throughout this period. Moreover, in all patients included in this study, the blood samples selected for proteomic analysis were drawn at a minimum time distance of 2 weeks from the last surgical



**Fig 2.** Discrimination of sera obtained from stage III melanoma patients in relation to disease progression within the first year of follow-up. Samples with recurrent disease were assigned 1, and no relapse was assigned 2. Mean values and associated t values ( $P < .05$ ) for 51 models are presented. Progressing patients (277 to 304) and nonprogressing patients (305 to 324) should have values less than 1.5 and greater than 1.5, respectively; 80% were correctly assigned. The concentration of S100-β ( $\mu\text{g/L}$ ) is given. Those exceeding 0.12  $\mu\text{g/L}$  are highlighted.

intervention. This distance was arbitrarily chosen and may be insufficient to rule out the possibility that soluble factors, which entered the peripheral blood due to the surgical procedure, significantly changed the serum proteomic expression profiles. Further studies are underway to answer these important questions.

Another feature of interest was that using an F score for testing equality of variances between two independent populations (stage I *v* stage IV melanoma serum samples) revealed that one region with mass values centered at approximately 11,700 Da was noted to be highly significantly different in variance between the two populations (refer to Fig 1B and Fig 1C). Initial data analysis by ANN suggested that the best predictive capability came from the 2,000- to 5,000-Da mass range, and little predictive value was obtained from the 10,000- to 15,000-Da mass range. This serves to underscore the fact that different statistical/bioinformatic strategies may well identify different molecular regions that may be useful as biomarkers (eg, those identified in Fig 1B that were generally associated with approximately 25% of stage IV patients) compared with regions that were useful in predicting stage-related disease as highlighted by the analysis of the ANN models. It is important, therefore, to consider that different data mining techniques may elicit different markers with differing importance depending on the context in which they may be utilized.

Recently, serum proteomic analysis has been shown to be able to discriminate between patients with ovarian cancer and unaffected individuals with 95% specificity.<sup>13</sup> Subsequently, two studies reported on the discrimination of healthy volunteers and individuals with benign prostate hyperplasia or prostate cancer (PCA) with 83% sensitivity and with up to 95% specificity.<sup>3,14</sup> Petricoin et al reported on the classification of 26% of sera from benign prostate hyperplasia wrongly as PCA; however, 10% of patients followed for 5 years developed PCA subsequently.<sup>14</sup> Further identification of predictive biomarkers of disease progression may therefore provide an important diagnostic tool. This study was performed using CIPHERGEN H4 protein chip technology (C16 hydrocarbon chemistry), which is no longer available. Similar chemistries (eg, H50) and alternate approaches (eg, ZipTip C18 hydrocarbon columns) may well provide enhanced methodologies for the identification of robust biomarker patterns.

Analysis of proteomic expression patterns may be used in the future to aid clinicians in selecting patients with poor prognosis for adjuvant therapeutic interventions or closer follow-up. The validation of this approach, taking into account biologic variability, differences in mass spectrometric analysers (eg, instrument resolution, mass accuracy, sensitivity), and statistical analysis, will be necessary before the methodology would be robust enough to be translated into a clinical setting. A recent report by Baggerly et al<sup>15</sup> has shown the inherent problems that can be associated with

the mining of MALDI-ToF data for biologic signatures in relation to some of the variables outlined above. To overcome these issues, multiple considerations will have to be taken into account if a systematically reproducible approach is to be found using biomarker expression patterns as predictive indicators. Considerations will have to focus on the problems of standardization of sample collection, storage, and processing. To achieve this goal, we have now implemented the use of relational clinical databases that will not only enable clinical data (eg, treatment regimen) to be correlated with expression profiling but will also serve to act as an audit system for how samples have been stored. For example, our tracking system will enable information to be obtained in relation to how a sample may have been manipulated during their storage (eg, how many times a sample was sectioned at  $-20^{\circ}\text{C}$  to determine whether temperature fluctuations from  $-80^{\circ}\text{C}$  to  $-20^{\circ}\text{C}$  may have an effect on biomarker stability). Standardizing material collection between multiple centers will be a fundamental requirement if new predictive tools are to be developed using these types of methodologies. Variation in these procedures may well lead to assays that are not stringent enough to work on a day-to-day basis between multiple centers (ie, they may well work well for a particular laboratory, but they are not robust enough to be applied to multiple laboratories). Furthermore, the equipment employed in this study uses linear ToF mass spectrometry in conjunction with a single dimension (reverse-phase chemistry) to deconvolute the serum proteome.

Given the inherent problems associated with low-resolution instruments to separate individual peaks during mass spectrometric analysis, combined with the fact that a single dimension of sample deconvolution may be unable to consistently provide the resolution necessary to identify molecules of importance. To this end, it will be possible to examine mass spectra derived from serum samples that will be deconvoluted in a number of ways in order to reduce dynamic range (eg, use of lectin columns, implementation of robotics to ensure consistent sample preparation); developing protocols analogous to 2D liquid chromatography that use off chip (not directly involving the use of SELDI-TOF chip clean-up) sample deconvolution protocols such as strong cationic exchange columns in the first dimension, followed by reverse phase in the second. Future studies will utilize high-resolution instruments (eg, resolution at full width half maximum  $> 20,000$ ) and postsource decay to enable sequence information to be obtained from mass values that seem to be important following data analysis and validate the use of data mining methodologies<sup>16</sup> in relation to peak normalization, background subtraction, and threshold values as part of the development process.

---

### **Authors' Disclosures of Potential Conflicts of Interest**

The authors indicated no potential conflicts of interest.

## REFERENCES

1. Anderson NL, Anderson NG: The human plasma proteome: History, character, and diagnostic prospects. *Mol Cell Proteomics* 1:845-867, 2002
2. Liotta LA, Ferrari M, Petricoin E: Clinical proteomics: Written in blood. *Nature* 425:905, 2003
3. Adam BL, Qu Y, Davis JW, et al: 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
4. Li J, Zhang Z, Rosenzweig J, et al: Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* 48:1296-1304, 2002
5. Mian S, Ball G, Hornbuckle J, et al: A prototype methodology combining surface-enhanced laser desorption/ionization protein chip technology and artificial neural network algorithms to predict the chemoresponsiveness of breast cancer cell lines exposed to paclitaxel and doxorubicin under in vitro conditions. *Proteomics* 3:1725-1737, 2003
6. Balch CM, Buzaid AC, Soong SJ, et al: Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *J Clin Oncol* 19:3635-3648, 2001
7. Bosserhoff AK, Kaufmann M, Kaluza B, et al: Melanoma-inhibitory activity, a novel serum marker for progression of malignant melanoma. *Cancer Res* 57:3149-3153, 1997
8. Ugurel S, Rappel G, Tilgen W, et al: Increased serum concentration of angiogenic factors in malignant melanoma patients correlates with tumor progression and survival. *J Clin Oncol* 19:577-583, 2001
9. Reibmann V, Ugurel S, Tilgen W, et al: Soluble HLA-DR is a potent predictive indicator of disease progression in serum from early-stage melanoma patients. *Int J Cancer* 100:580-585, 2002
10. Deichmann M, Brenner A, Bock M, et al: S100-beta, melanoma-inhibiting activity and lactate dehydrogenase discriminate progressive from non-progressive American Joint Committee on Cancer stage IV melanoma. *J Clin Oncol* 17:1891-1896, 1999
11. Ghanem G, Loir B, Morandini R, et al: On the release and half-life of S100B protein in the peripheral blood of melanoma patients. *Int J Cancer* 94:586-590, 2001
12. Acland K, Evans AV, Abraha H, et al: Serum S100 concentrations are not useful in predicting micrometastatic disease in cutaneous malignant melanoma. *Br J Dermatol* 146:832-835, 2002
13. Petricoin EF, Ardekani AM, Hitt BA, et al: Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359:572-577, 2002
14. Petricoin EF 3rd, Ornstein DK, Paweletz CP, et al: Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 94:1576-1578, 2002
15. Baggerly KA, Morris JS, Coombes KR: Reproducibility of SELDI-ToF protein patterns in serum: Comparing data sets from different experiments. *Bioinformatics* doi: 10.1093/bioinformatics/btg484
16. Ball G, Mian S, Holding F, et al: An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics* 18:395-404, 2002