

Serum protein expression profiling for cancer detection: Validation of a SELDI-based approach for prostate cancer

William E. Grizzle^{a,*}, Bao-Ling Adam^b, William L. Bigbee^c, Thomas P. Conrads^d, Christopher Carroll^e, Ziding Feng^f, Elzbieta Izbiicka^g, Moncef Jendoubi^h, Donald Johnseyⁱ, Jacob Kaganⁱ, Robin J. Leach^e, Diane B. McCarthy^j, O. John Semmes^b, Shiv Srivastava^k, Sudhir Srivastavaⁱ, Ian M. Thompson^e, Mark D. Thornquist^f, Mukesh Vermaⁱ, Zhen Zhang^l and Zhiqiang Zou^k

^aUniversity of Alabama at Birmingham, Birmingham, AL, USA

^bEastern Virginia Medical School, Norfolk, VA, USA

^cUniversity of Pittsburgh Cancer Institute, Pittsburgh, PA, USA

^dSAIC-Frederick, Frederick, MD, USA

^eUniversity of Texas Health Sciences Center San Antonio, San Antonio, TX, USA

^fFred Hutchinson Cancer Research Center, Seattle, WA, USA

^gInstitute of Drug Development, San Antonio Cancer Institute, San Antonio, TX, USA

^hMilagen, Inc., Richmond, CA, USA

ⁱNational Cancer Institute, Rockville, MD, USA

^jCiphergen Biosystems, Inc., Fremont, CA, USA

^kUniformed Services University of Health Sciences, Rockville, MD, USA

^lJohns Hopkins Medical Institutions, Baltimore, MD, USA

Abstract. Multiple studies have reported that analysis of serum and other bodily fluids using surface enhanced laser desorption/ionization time of flight mass spectrometry (SELDI-TOF-MS) can identify a “fingerprint” or “signature” of spectral peaks that can separate patients with a specific disease from normal control patients. Ultimately, classification by SELDI-TOF-MS relies on spectral differences in position and amplitude of resolved peaks. Since the reproducibility of quantitation, resolution and mass accuracy of the SELDI-TOF-MS, or any high throughput mass spectrometric technique, has never been determined this method has come under some skepticism as to its clinical usefulness. This manuscript describes a detailed design of a three-phase study to validate the clinical usefulness of SELDI-TOF-MS in the identification of patients with prostatic adenocarcinoma (PCA). At the end of this validation study, the usefulness of the general SELDI-TOF-MS approach to identifying patients with PCA will be demonstrated and how it compares with PCA diagnosis by measuring prostate specific antigen.

Keywords: SELDI-TOF-MS, prostate adenocarcinoma, validation

1. Introduction

Prostatic adenocarcinoma (PCA) is diagnosed in about 190,000 men each year and is responsible for 30,000 deaths annually [2]. As such, it is the second major cause of cancer deaths in men. When PCA is lo-

*Corresponding author: William E. Grizzle, M.D., Ph.D., University of Alabama at Birmingham, Zeigler Research Building-Room 408, 703 South 19th Street, Birmingham, AL 35233-0007, USA. Tel.: +1 205 934 4214; Fax: +1 205 975 7128; E-mail: grizzle@path.uab.edu.

calized to the prostate, it can be cured in most cases by local therapy including radical prostatectomy. Therefore, early detection of prostate cancer is very important to the management of this disease process.

Since its discovery in 1971 by Hara et al. [16], screening with prostate specific antigen (PSA) in serum has become the most valuable approach for the early detection of PCA. Several large clinical studies have demonstrated that PSA alone is superior to digital rectal examination (DRE) or any other existing screening method for PCA, and that PSA combined with DRE is the most effective early detection modality for prostate cancer [4,7,11,17,18]. With the advent of widespread PSA screening programs, a significant trend toward diagnosis of earlier stage PCA has occurred. This sizable shift in stage-specific incidence rates toward earlier stage disease has been accompanied by concurrent declines in new cases diagnosed with poorly differentiated tumors and a decrease in patients diagnosed with metastatic disease [15,21,26]. This combination of events has suggested that PSA testing may have had a beneficial impact on mortality [28].

The positive predictive value of PSA has been reported to be ~10% in men with serum PSA levels of <4 ng/ml and >80% for men with levels >20 ng/ml. While most patients with PSA levels <10 ng/ml have early stage disease, over 50% of patients with PSA levels >10 ng/ml are found to have advanced disease [17]. These results emphasize the need to use a low PSA cut-off value for the early detection of prostate cancer. Unfortunately, a detection threshold value of 4 ng/ml leads to a false positive rate of ~65% in typically screened populations [3]. This high false positive rate results because PSA is more *prostate specific* than *PCA specific*, as elevated PSA levels are found in benign prostatic hyperplasia (BPH), prostatic intraepithelial neoplasia (PIN), prostatic ischemia/infarction, and acute and chronic prostatitis [5]. Thus, in men with a PSA between 4–10 ng/ml, many prostate biopsies reveal only benign conditions. In contrast, approximately 20% of PCA patients have PSA levels <4 ng/ml at diagnosis [4, 8]. Thus, in spite of the overall success of PSA screening, there continues to be a need for more sensitive and specific methods for early detection of PCA either independent of or as an adjunct to PSA screening.

Two studies have demonstrated that surface enhanced laser desorption/ionization (SELDI) time of flight mass spectroscopy (TOF-MS) can be used to generate serum protein expression profiles in patients with PCA that can be discriminated from those of healthy controls using pattern matching algorithms ap-

plied to this multidimensional proteomic data [1,23]. In these initial studies, the SELDI-TOF-MS method coupled with a pattern-matching algorithm has identified PCA with a sensitivity of 83% and specificity of 97%. Subsequently, a boosted decision tree analysis increased both sensitivity and specificity to 97%. These results were judged by the NCI Early Detection Network (EDRN) to be encouraging enough to warrant validation of the SELDI-TOF-MS method for the early detection of PCA.

SELDI-TOF-MS is a novel technique for determining differential protein expression and is useful in biomarker discovery [22] due to its unique ability to detect analytes at trace (low femtomole) levels in complex biological samples, e.g., serum, urine, CSF, tissue culture extracts, and tissue lysates. The ProteinChip® Arrays used in this process fall into two broad categories of chromatographic and pre-activated surfaces. For protein expression profiling, chromatographic surfaces (e.g. anionic, cationic, metal affinity, and reverse phase) are used. Crude samples such as serum can be applied directly to the array and proteins are retained based on their biochemical properties. Washing of the array removes nonspecifically bound proteins, salts, and other interfering compounds, greatly reducing signal suppression. Therefore, unlike many other types of protein arrays, the ProteinChip® Arrays do not require a prior knowledge of the proteins of interest. Investigators can custom make more conventional protein arrays using the pre-activated ProteinChip® Arrays. This technology allows the user to covalently attach a specific capture molecule to the biochip to create a customized affinity array. The capture molecules can include antibodies, small molecules, peptides, nucleotides, and aptamers. Alternative forms of mass spectrometric analysis generally relies on front-end fractionation or sample purification prior to loading on a target or probe often resulting in the loss of low abundance proteins.

The ProteinChip Reader is a compact TOF-MS instrument specifically developed for SELDI-TOF-MS applications. After capture of the sample on the array, an energy absorbing matrix (EAM) solution, that enhances laser energy transfer and analyte ionization, is added to promote laser-based desorption and ionization. This part of the process is similar to MALDI but has the advantage provided by the chemistry of the solid support surfaces of the ProteinChip. Retained proteins are then detected by the ProteinChip Reader. Proteins and peptides with molecular weights ranging from less than 1000 to 200,000 daltons (Da) can be detected.

The time required for the discovery of molecular markers for the early detection of specific neoplastic

processes is unpredictable. However, once a potentially useful biomarker is identified, a number of criteria must be met before it becomes available clinically. In addition to standard performance criteria such as sensitivity and specificity, the assay technique must be standardized and then the marker compared to the current 'gold standard' in a rigorous validation process. Pepe et al. [20] have published a general pathway for the identification of markers of early cancer detection including a validation step [20]. This article provides the details of the validation of markers prior to the clinical use of SELDI-TOF-MS for diagnostic and therapeutic interventions. This detailed validation process, focused on PCA, was developed during a meeting of a subgroup of the Early Detection Research Network Genitourinary (EDRN GU) Collaborative Group. As an initial product of this Collaborative Group, we present a proposed validation approach that is not only specific for the early detection of PCA but which also can be generalized to the validation of molecular approaches to early detection for other forms of neoplasia, either by SELDI-TOF-MS or by other screening methods.

2. Study design, design considerations and methods

The validation process developed for the early detection of PCA using SELDI-TOF-MS serum expression profiling is divided into three stages. In Stage I, the validation procedure focuses on the reproducibility and portability of the method and the standardization of the methodology at seven clinical – laboratory sites. These sites include Eastern Virginia Medical School (EVMS), Johns Hopkins University Medical Institutes (JHU), Uniformed Services University of Health Sciences (USUHS), University of Alabama at Birmingham (UAB) Medical Center, University of Pittsburgh Cancer Institute (UPCI), National Cancer Institute Frederick Cancer Research Facility (NCIFCRF), and University of Texas Health Science Center San Antonio (UTHSCSA). Stage II is the demonstration that PCA can be detected with a high sensitivity and specificity by multiple laboratories using first, serum samples from a single-institution biorepository-based study cohort and second, in an independent PCA patient and control population. Finally, in Stage III, a well-defined patient population such as the Prostate Cancer Prevention Trial (PCPT) will be utilized to ensure the sensitivity and specificity of the method in a population of clinically well-defined PCA cases and prostate biopsy proven

negative controls. This study will be combined with an analysis of the predictive power of the method in prospective serum samples from emergent PCA cases in the PCPT cohort obtained years prior to clinical diagnosis.

2.1. Stage I. Reproducibility and portability

There are three sequential components of Stage I that are designed to demonstrate that comparable SELDI-TOF-MS serum protein expression profiling results can be obtained in multiple laboratories. In developing a detailed validation study for the early detection of PCA using SELDI-TOF-MS, the Working Group first considered reproducibility and portability. That is, the identical methodology will be established at multiple sites with the goal of obtaining the same results at each site. Ultimately it must be demonstrated that multiple sites using the same instrumentation and protocols consistently obtain comparable data on a set of well-characterized subject samples. For SELDI-TOF-MS, these data would include the same diagnostic peaks that have been identified by the EVMS group (reference site) to separate PCA cases from non-PCA controls.

To optimize reproducibility and portability, the instruments used in this analysis at each site will be calibrated with EVMS and each site must reproduce the results from a set of samples previously analyzed by EVMS. For the six analytical sites plus the EVMS reference site, this testing will be completed within six months; at that time, all data will be analyzed by the Data Management and Coordinating Center (DMCC) of the EDRN. After six months, the Working Group will meet again and evaluate the results of the portability and reproducibility portion of the validation study (i.e., Stages IA-C).

2.2. Stage IA: Standardization of SELDI-TOF-MS instrumentation and laboratory protocols

Six of the laboratories participating in the validation study (EVMS, JHU, USUHS, UAB, UPCI, NCIFCRF, and UTHSCSA) will utilize the same instrumentation. This required analytical equipment is Ciphergen Biosystems Inc. SELDI-TOF-MS, PBSII Reader with Biomarker Patterns™ Software and a Biomek® 2000 Sample Preparation Robotic System from Beckman Instrumentation with the Ciphergen Biomarker System Integration Package. NCIFCRF will use a different system.

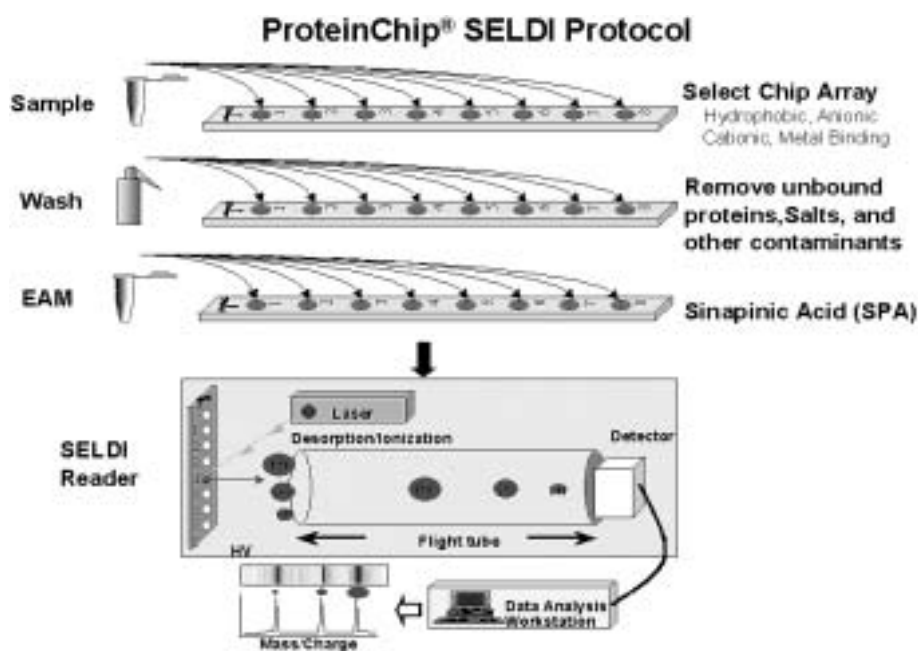


Fig. 1. Illustration of SELDI Time-Of-Flight (TOF) Mass Spectrometry. (Modified with permission from Ciphergen Biosystems).

The SELDI-TOF-MS systems within the seven participating laboratories will be calibrated by Ciphergen engineers and scientists using insulin to test the resolution and IgG to test the sensitivity of the instrument. The instruments will then be standardized utilizing a quality control assay developed by EVMS. This quality control assay consists of a pooled quality control serum sample applied to an IMAC3-Cu ProteinChip® array prepared using standardized protocols. These chips will be prepared and loaded with samples at EVMS and standard spectra from these chips will be obtained at EVMS before standardizing the other laboratories. EVMS investigators and staff also will visit each of the sites with the Ciphergen personnel to oversee calibration and standardization of each of the SELDI-TOF-MS instruments.

Normalization of instruments across multiple sites requires specific hardware and software components. In the process of standardizing the instruments on the quality control serum sample, instrument settings such as detector voltage, laser intensity, and sensitivity will be adjusted. All spectral data will be collected using the same lag time and number of laser shots. After data collection, each spectrum will be calibrated for mass using the same peptide calibrants. Spectra will be normalized using total ion current normalization with the same normalization coefficient and low mass cutoff. All spectra will be processed using the same baseline

subtraction protocol and peaks will be detected using a uniform definition of requisite signal-to-noise ratio and mass window.

In this stage, EVMS personnel will run the previously prepared quality control (QC) chips, together with new chips prepared at each site loaded with the QC serum standard and adjust instrumentation so that the profiles are equivalent to those obtained at EVMS. We expect to see four major peaks (5,908, 7,768 and 9,293 Da)-with resolution above 450 and intensity in the range of 20 to 40 [1]. (These are the standards for laser energy, detector sensitivity and detector voltage adjustment.) After instrument adjustment, two bioprocessors of QC samples (24 protein chips with 8 spots per chip) will be processed and analyzed using SELDI-TOF-MS to confirm the quality of the assay. For each of the following three months, one bioprocessor of QC samples (12 protein chips) will be processed and read using SELDI-TOF-MS (12 chips per month) to monitor temporal reproducibility. In addition, one chip of QC and peptide standards will be run every Monday morning to monitor instrument performance.

2.3. Stage 1B: Replicate analyses of EVMS sera from PCA patients and controls

The primary purpose of Stage 1B will be to demonstrate that each participating laboratory can obtain

consistent SELDI-TOF-MS results to discriminate between the sera of PCA and non-PCA subjects. Following the Stage IA standardization of the SELDI-TOF-MS system at a site, the site will then receive 14 PCA and 14 non-PCA sera from EVMS. These samples will be blinded and have been previously analyzed at EVMS. Samples will be selected from a previous study at EVMS for which the non-PCA subjects are age-matched normals (defined as subjects with PSA <4 and normal DRE), and the PCA subjects are patients who have undergone radical prostatectomy. Criteria used to select these 14 samples out of 186 PCA subjects and 219 age-matched normals include 1) spectra show no signs of protein degradation and 2) at least 1 ml of serum remaining in the serum bank. Samples will be analyzed blindly by each of the five laboratories as well as EVMS. Each of the six sites will prepare these samples on IMAC3-Cu chips using their Biomek[®] robotic system and analyzed on the SELDI-TOF-MS system. All sites will send their primary data and processed spectra to the Co-Investigators at the DMCC who will analyze each site's ability to identify the peaks required for accurate analysis and classification of these samples.

The criteria for agreement among labs are based on the agreement in both the resolution and intensity of the diagnostic peaks. Since the true protein mass is unknown at each of the protein mass peaks detected by SELDI-TOF-MS (there is always some measurement error, no matter how small it is), the resolution agreement is determined by the agreement of the locations of peaks as measured at each of the six other analytical sites and EVMS and between each site and the other six laboratories combined. More specifically, after each laboratory performs peak identification and alignment, the coefficient of variation (CV) and intra-class correlation coefficient (ICC) for measures of peak location and of peak intensity will be calculated for each peak. The between-laboratory variation will be compared to the between-sample variation. The direct assessment of the between-laboratory agreement will apply the classifier previously developed from the EVMS data (the 28 samples are part of the data used to develop the original classifier and thus are classified correctly) to the data from each laboratory. The same synchronization techniques applied to the pooled QC sample in Phase IA will be used to improve the inter-laboratory agreement.

2.4. Stage IC: Analyses of sera from PCA patients and controls drawn from each participating site

Stage IC will be performed at each of the seven sites following the successful completion of Stage IB. The

objective of Stage IC is to demonstrate that diversely collected, processed and stored sera of PCA and non-PCA subjects from the six sites can be prepared and analyzed accurately by each of the sites using SELDI-TOF-MS. This stage will also identify instabilities that may occur subsequently in the instrumentation used in the study as well as subtleties in the collection, processing, and storage of blood samples that may impact the profiling results.

A total of 35 cases of PCA and 35 controls without PCA will be analyzed blindly by each of the six participating laboratories using the same protocols developed in Stages IA and IB. The rest of this stage will follow the protocols described for Stage IB with analysis of results being performed by the DMCC. Serum samples from these 70 subjects will be separated into six 0.5 ml aliquots and stored within 4 hours of collection at -70°C or colder until provided to the sites. Five cases and five controls will be collected from each site. The controls will be age-matched normals with PSA <2.5 ng/ml and normal DRE. (Preferably, control subjects will also have had a negative prostate biopsy.) Cases will include pre-treatment samples from prostatectomy patients.

Similar to Stage 1B, the CV and ICC will be calculated for measures of peak location and of peak intensity. The existing classifier built on the data from EVMS may not correctly classify all 70 samples from different laboratories. Therefore, agreement will be based on whether data for these 70 samples from each of other 6 laboratories are classified the same as the data from EVMS. This agreement is reasonable at this stage since the goal of Stage I is to assess the impact of the between laboratory variation on the consistency of diagnosis/detection.

3. Stage II. Validation and refinement of the predictive algorithm in an independent, well-defined PCA case-control series

Stage II of the validation study is designed to demonstrate that the method accurately predicts presence or absence of PCA in an independent, large case-control series, collected by multiple sites and analyzed at two independent but standardized sites (EVMS and UAB). Stage II of the validation study will also demonstrate that the algorithm developed by EVMS and DMCC [1, 23] to analyze SELDI-TOF-MS data can be used on profiling data obtained from a different group of PCA patients and controls analyzed at EVMS, the EDRN

Discovery Laboratory, and UAB, an EDRN Biomarkers Validation Laboratory. After the current algorithm is tested and validated, the PCA cases and controls will be randomly assigned to a learning set and a new optimized algorithm will be developed. Subsequently, this algorithm will be validated in a test set consisting of the remainder of the specimens. Thirty percent of the total samples, i.e., 75 for each group, will be used as the test set.

The first step of Stage II will be to estimate the sample size required. This calculation will depend on prior measurements at the EVMS Developmental Laboratory including the initial determination of sensitivity and specificity during developmental studies. Two groups of 250 patients each will comprise two populations – one low risk PCA group with clinical T1-T2 stage disease, and a Gleason Score of 6 or less and a high risk PCA group for clinically aggressive disease with T1-T2 stage and a Gleason Score ≥ 7 . This Gleason cut-point has been selected due to the clear difference in biologic activity between the <7 and ≥ 7 groups [12]. The study is limited to T1-T2 disease so that only potentially curable cases will be studied. A total of 250 controls (non-PCA) will be selected, collected and their blood samples processed and stored identically to the patients' samples. The controls will have PSA values ≤ 10 ng/ml, a normal DRE, and negative biopsies (sextant minimum) with no PIN and no proliferative inflammatory atrophy (PIA) identified at biopsy.

The primary goal of Stage II is to validate the existing classifier developed from EVMS data on new samples of serum from PCA patients or normal controls with well defined inclusion/exclusion criteria, and from diverse laboratories and clinical populations. If the existing classifier works well, the study will proceed to Stage III even though a better classifier may be developed as a secondary aim of Stage II. If the existing classifier does not work well on these diverse samples, the first 125 samples from each group will be used as a training sample to modify the classifier and the second 125 samples from each group will be used as a new test sample to assess the performance of the new classifier. The existing classifier has both a sensitivity and specificity of 97%. If it is assumed that both the sensitivity and specificity of the classifier on the new samples will be 92%, with 125 samples per group, the study has 86% power to rule out an unacceptable sensitivity or specificity of 81% using a joint 95% confidence region for sensitivity and specificity. When testing the existing classifier on all samples (250 per group), the study has 91% power to rule out an unacceptable sensitivity

or specificity of 84%. This sample size calculation is based on the potential clinical use of SELDI-TOF-MS protein profiling as a complementary test to PSA. If a patient has an elevated PSA and if a test with both high specificity and sensitivity is negative, the patient would not be biopsied. If the classifier shows very high specificity (i.e., 98%) and good sensitivity, it then has potential to be used as an independent population screening tool to replace PSA. In that case, a much larger prospective study [19] is required to confirm the high specificity. At the current level of validation, such a large investigation is not warranted.

Serum samples will be obtained prior to biopsy from the patient population identified by the Working Group. This population consists of 250 controls, 250 patients with high risk PCA and 250 patients with low risk PCA. All samples will be processed, aliquoted, and frozen and stored at -70°C or colder within four hours of collection. Each sample will be stored as multiple 0.5 ml aliquots. For samples collected prior to the design of this study, requirements will be less restrictive: storage at -70°C or colder, not thawed prior to this study, and at least 1 ml of sample available to support the validation study.

An interim analysis of the Stage II validation study will be performed when the sample size reaches 125 controls, 125 low risk PCA and 125 high risk PCA cases. Depending on the efficiency of identifying and obtaining these samples, this point may be reached by September 2003. If the study is favorable at the interim analysis, the Working Group will request samples for the described Stage III portion of the validation study from the Prostate Cancer Prevention Trial (PCPT) [10].

4. Stage III. Validation in serum samples from subjects enrolled in the Prostate Cancer Prevention Trial (PCPT)

The primary purpose of Stage III of the validation study will be to validate definitively the early detection of PCA using SELDI-TOF-MS serum expression profiling. This final stage of the validation study was designed using samples from the PCPT based on the criteria described subsequently. A comprehensive protocol will be presented to the PCPT Steering Committee for approval, based upon Stages I and II of this study. This study will require the use of a pristine control group that includes subjects who have an exceptionally low likelihood of occult PCA. This parameter is ensured by the analysis of initial samples of sera drawn when

patients first entered the PCPT trial, plus the clinical determination that subjects do not have PCA by a negative biopsy (minimum sextant biopsy) including no PIN and no PIA, at the end of the trial (up to seven years after the initial blood sample.) The details of the PCPT have been reported previously [10].

This study will use serum samples from patients with early PCA as determined by prior PSA measurements and by biopsy (T1 and T2 disease by at least sextant biopsies). The use of the PCPT samples will permit comparison of temporally distant detection of the development of PCA by PSA and SELDI-TOF-MS analyses. For example, the study will evaluate PSA, free PSA and protein expression profiles by SELDI-TOF-MS in serum samples collected up to seven years before the tumor was detected clinically.

The study will validate the detection of PCA in samples of serum collected prior to PCA diagnosis from subjects in both the placebo and finasteride treatment arms of the PCPT. Because finasteride may be used to treat a large number of patients with BPH in the future, the Working Group concluded that it is necessary to test SELDI-TOF-MS for PCA detection in such a population. The validation will use a sample collection obtained at a variety of sites under typical clinical conditions. The variation of sample collection parameters should increase the "noise" in the analytical system and may more accurately reflect the general sampling conditions to be encountered in typical clinical settings.

The final set of samples requested from the PCPT as well as aspects of the design of Stage III will depend upon the outcomes of Stages I and II of this validation study. First, the Working Group will evaluate whether or not the stratification of patients into low risk PCA and high risk PCA is appropriate. Similarly, we will evaluate racial and ethnic subgroups to establish the impact of these variables. Also, the exact number of patients to be included in the analysis will be determined based on the analysis of the potential α and β errors based on the Stage II results.

Patients with PCA: These patients must have no hormonal therapy or pre-treatment prior to obtaining samples of sera. The cases will be limited to T1 and T2 disease, i.e., no evidence of metastases or extra-prostatic disease: The definition of low risk PCA is patients with tumors with Gleason Score =6; similarly, high risk PCA, Gleason Score ≥ 7 . The segregation into low risk and high risk groups of PCA was selected by the Working Group to test the hypothesis that SELDI-TOF-MS can discriminate low from high risk disease. To our knowledge, this study is the first attempt to develop a

biomarker for prostate cancer that takes into account the range of biologic activity of the disease.

Controls: Controls are defined as: subjects with a PSA level of ≤ 10 ng/ml, normal DRE not suspicious for malignancy, negative biopsy (sextant biopsy minimum), with no PIN or PIA, and no prior hormonal therapy. The PSA range of 4–10 ng/ml for controls was included so that patients with BPH and an increased PSA would not be excluded from the study. (By excluding this group, SELDI-TOF-MS might indeed become a marker more for the absence of BPH than the presence of PCA.) However, samples will be stratified in the <4 ng/ml and 4–10 ng/ml ranges and analyzed separately.

Stage III of the validation study is designed to use a highly selected group of men who were likely either not to have PCA or to have localized prostatic adenocarcinoma at the time of sampling. Sample sets, in addition to the PCPT (e.g., the Prostate, Lung, Colorectal and Ovary [PLCO] and other cohorts) were reviewed. Because of the extremely well characterized nature of PCPT as well as its PSA-independent end-of-study prostate biopsy, the PCPT population was deemed to be the optimal validation study sample. The Working group identified further opportunities to include smaller serum banks at participating institutions where samples of sera from a defined cohort of PCA patients are available and linked to comprehensive clinico-pathologic data including data from whole-mounted radical prostatectomy specimens and longitudinal follow-up after radical prostatectomy. Such serum specimens will be especially useful for defining serum protein expression profiles related to extent of disease, response to treatment and biologic activity including clinical outcomes such as PCA recurrence, clinical metastasis and death.

Unique advantages of using the PCPT cohort includes:

- a) The availability of controls in the PCPT cohort who are unlikely to harbor significant PCA at the time of their initial blood sampling: seven-year, end-of-study negative biopsy with normal PSA and normal DRE for the duration of the study.
- b) PCA cases identified in the PCPT with multiple serum samples available at diagnosis of PCA and sequential samples for several years before PCA was diagnosed.
- c) Subjects treated with finasteride will provide important information on the effects of this agent on SELDI-TOF-MS serum protein expression profiles.

Our approach to the design of this study expands the analysis of early detection biomarkers proposed by Pepe et al. [20] by proposing a specific model for early detection validation studies in general and SELDI-TOF-MS studies specifically. In this validation study, additional SELDI-TOF-MS profiling data will be collected and analyzed using existing predictive algorithms. As an integral part of the experimental design, new potentially more accurate algorithms using learning and testing sets will be developed and validated. While the test for PSA is quite sensitive, it is not specific for PCA. Specificity is especially a problem in the abnormal range of 4–10 ng/ml in which BPH is the most frequent confounding disease. To increase the specificity of PSA, assays for free PSA, PSA velocity, PSA density and new forms of pre-PSA have been or are undergoing development for the early detection of PCA. In addition, because patients who have PSA values of <4 ng/ml are biopsied infrequently, we do not know how many cases with Gleason scores of 7–10 are being missed in men with PSA levels below this upper limit of the normal range.

5. Discussion and summary

The discovery and clinical validation of markers of early detection of a wide range of neoplastic processes is the primary goal of the Early Detection Research Network (<http://www.cancer.gov/edrn>). As new biomarkers are discovered through proteomic approaches, the necessity to validate and ultimately use them in a clinical setting increases. This can only be done as a collaborative effort among research communities. The National Cancer Institute (USA) has taken a leading role in this regard by creating the Early Detection Research Network (EDRN). Since its inception in 2000, EDRN has supported collaborative research on molecular, genetic, and other biomarkers related to the early detection of cancer and to risk assessment. The EDRN initially was comprised of eighteen Biomarker Development Laboratories (BDLs), nine Clinical and Epidemiology Centers (CECs), three Biomarker Validation Laboratories (BVLs), and a central Data Management and Coordinating Center (DMCC). Subsequently, 12 associate member laboratories have been added to the EDRN.

The mission of the EDRN is to streamline the development, testing, and evaluation of promising biomarkers and technologies to minimize the time from initial exploratory studies to the dissemination

of a validated biomarker for clinical use. This network brings together national and international experts, from academia and industry, to promote biomarker discovery, validation and help translation into clinical practice. Exciting information is expected to emerge from such a collaborative effort that can ultimately be applied to population screening for the early detection/diagnosis of cancer.

Pepe et al. [20] proposed five phases in the development of biomarkers for the early detection of cancer. These five phases are 1) preclinical exploratory studies, 2) clinical assay development for clinically established disease, 3) retrospective longitudinal repository studies, 4) prospective screening studies, and 5) cancer control studies. The three stages of this proposed study reflect the special challenges of using protein profiling as a “biomarker” for the early detection of prostate cancer. For example, Phase I ranks and prioritizes biomarker candidates and selects the most promising biomarkers for Phase II studies. However, if a study directly validates an existing and tested protein profile, as in this case, then Phase I is not applicable.

Because validation of SELDI-TOF-MS serum protein expression profiles as a method to detect preclinical PCA is somewhat different, the term “stage” in the steps of validation is used to prevent confusion with the 5 “phases” of biomarker development. Specifically, stage 1 of this study focuses on portability and standardization at other laboratories; if the biomarker is a known protein or gene, portability and standardization is not addressed directly until later phases (e.g., 4 and 5). However, it is crucial to establish portability and reproducibility at an early stage when the “biomarker” is a protein profile consisting of unidentified proteins. After Stage 1 of our validation study is completed, Stage 2 probably best matches to independently repeating the Phase 2 study that already has been completed successfully at the BDL at EVMS [1,23]. Independently repeating a Phase 2 study is sometime necessary if a Phase 3 study would be very expensive to conduct. Since protein profiling uses high dimensional protein data, the danger of bias due to systematic differences in cancer and control specimens from one site may be more pronounced because multiple proteins may be affected by collection and processing procedures and unique aspects of the patient population; these biases alone may contribute to an erroneous discrimination between cancer and control samples. On the other hand, if the biomarker is an identified single molecule, differences in collection and processing affects on the evaluation of this molecule and any bias this might introduce should

be identified more easily. It also is more important to repeat a Phase II test using samples from multiple sites when Phase I is by-passed.

Biomarker development for the early detection of PCA is challenging as 1) there is a risk of detection of indolent disease and 2) as definition of 'controls' is uncertain due to the risk of undiagnosed PCA in men with a negative prostate biopsy. Thus, an important objective of this study is to detect aggressive subtypes of prostate cancer; this was not an objective of the previous EVMS Phase II study. Therefore, a new phase II study with patient selection criteria satisfying the objective of this study is necessary.

Even though Stage II of this study is classed as a Phase II study, it is actually beyond a Phase II study because most of the patients "with cancer" were not identified symptomatically but rather by the very sensitive screening test, PSA. Also, all cancer patients included in Stage II of this study have T1 or T2 disease so that they most likely would be cured by surgery because only 10 to 15% of T1 or T2 disease recurs after radical prostatectomy. Similarly, the controls for the Stage II testing have PSA values ≤ 10 ng/ml, a normal DRE, and negative biopsies (sextant minimum) with no PIN or PIA. Thus, Stage II of this study clearly will identify patients with PCA who can be cured by surgery so this stage at least meets some of the goals defined by Pepe et al. for a Phase III study.

Stage 3 using PCPT samples would be theoretically a phase 3 study. However, it is much more than a simple phase 3 study because the collection of samples in the PCPT was designed as a prospective collection, and the PCPT had clearly-defined clinical outcomes, including cases and controls that would be challenging to obtain in any future study. Also, biopsies of the prostate in the PCPT study were prompted by a sensitive screening biomarker, PSA. Thus, completing Stage 3 will be tantamount to conducting a prospective screening program or to completing some of a Phase 4 study, "Prospective Screening Studies"; however, the results of this SELDI-TOF-MS study do not prompt a biopsy/therapeutic outcome as required by the definition of a Phase IV test.

If Stage 3 is completed successfully with a true positive rate of greater than 90% and a false positive rate of less than 50%, it should warrant consideration for approval of SELDI-TOF-MS as a validated diagnostic tool. A further multi-institutional prospective trial to identify the false referral rate and the characteristics of the PCA detected by prospective profiling would be warranted. Such information will not be obtained from

the PCPT samples; but, the PCPT study has the potential to assess the lead time from a positive profile to a positive biopsy. In a prospective Phase IV study, if a profile were positive but the biopsy were negative, it would not be known until several years later if a patient has PCA or not. However, if the biopsy were positive in a Phase IV study and a radical prostatectomy were chosen as therapy, the pathologic and clinical stage of the cancer would be known; this is the same situation as for the PCPT samples as driven by PSA testing. The EDNRN does not plan to organize a Phase V study that, if necessary, best fits as a component of a very large clinical trial. Such studies might not be necessary before SELDI-TOF-MS profiling could be implemented in the general population for the early detection of PCA if SELDI-TOF-MS were shown to be a superior test to PSA both in sensitivity and specificity.

6. Future opportunities

While investigators have used a variety of different bioinformatic algorithms for pattern discovery, the most common analytical platform is comprised of a PBS-II SELDI-TOF-MS. Serum protein pattern analysis can also be conducted using a hybrid quadrupole time-of-flight (Qq-TOF) MS platform fitted with a SELDI ProteinChip array interface to test the hypothesis that higher resolution MS data will generate diagnostic models possessing higher sensitivities and specificities. The routine resolution obtained is in excess of 8000 (at $m/z = 1500$) for the Qq-TOF MS and only 150 (at $m/z = 1500$) for the PBS-II. Indeed, simulations demonstrate the ability of the Qq-TOF MS to completely resolve species differing in m/z of only 0.375 (e.g., at $m/z = 3000$) whereas complete resolution of species with the PBS-II TOF MS (routine resolution ~ 150) is only possible for species that differ by m/z of 20 (simulation not shown). While the higher resolution mass spectrometers are capable of peptide fragmentation and sequence analysis, it is unclear whether higher resolution is necessary to develop more sensitive and specific "fingerprint" assays for the early detection of cancer. Because Ciphergen's ProteinChip® Arrays used in the PBSII ProteinChip Reader also can be read using a ProteinChip Interface coupled to the QSTAR, the arrays analyzed in our study could be analyzed subsequently in QSTAR coupled systems to determine if higher resolution results in improved sensitivity and specificity or overall reproducibility in detec-

tion of PCA. NCIFCRF will participate in this aspect of the study.

Tandem MS instruments equipped with electrospray ionization (ESI) sources that analyze proteins in aqueous solution, can provide different but complementary information. ESI (as well as MALDI) sources are compatible with instruments that provide very high mass accuracy (i.e. <2 ppm), a level critical for accurate determination of peptide sequences. Analysis of the peptide MS/MS spectra followed by searching against the sequence database to identify the protein is a reliable method for the identification of specific small peptides, some of which form the peaks of interest in the SELDI-TOF-MS protein expression profiles of sera of patients with PCA. However, full application of ESI-MS instrumentation would require the introduction of significant front-end fractionation approaches using advanced microfluidics.

While substantial efforts of this study are focused on protein expression profiling by SELDI-TOF-MS, another goal is to identify serum proteins that are components of the protein peaks, distinguishing patients with PCA from healthy individuals. We can expect that after SELDI-TOF-MS profiling, aliquots of sera from the positive and negative cases will be available for further analysis. The first step toward protein identification will be a reduction of sample complexity by means of electrophoretic separation and/or chromatographic methods. The former may be similar to 2D PAGE, however, this and similar methods are biased toward larger and more abundant proteins. For low abundance serum prostate cancer biomarkers that may be present at parts per million levels and smaller molecular species, some form of affinity capture may be a preferred method. Other approaches for small molecules may use tandem MS-MS methods.

Although proteomic as well as genomic approaches are beginning to define common phenotypic alterations in PCA that may be valuable targets for the early detection of cancer, no serum marker thus far reported has greater potential than SELDI-TOF-MS serum protein profiling. With the introduction and perfection of emerging technologies that would be capable of assessing greater volume of the proteome it is anticipated that new defined serum biomarkers for both PCA diagnosis and prognosis will become available and the validation trial proposed here will serve as an excellent paradigm for these potential emerging PCA biomarkers. In addition, the design for validation proposed here for the detection of PCA using primarily SELDI-TOF-MS can be readily generalized to encompass the early detection

of other cancers using SELDI-TOF-MS. In addition, this model for the validation of the early detection and classification of prostate adenocarcinoma using protein expression profiles, as determined by SELDI-TOF-MS [19] could be adapted in the near future to validate other emerging proteomic technologies such as protein and antibody microarrays [6,13,14,24,25,27].

Acknowledgements

Supported by the National Cancer Institute Early Detection Research Network (EDRN). Other than the corresponding author, all other authors have contributed to this manuscript equally and are listed in alphabetical order. Elzbieta Izbicka: Supported by grants from the San Antonio Area Foundation from the Semp Russ Foundation, Shelby Rae Tengg Foundation, and P30 CA54174

References

- [1] B.-L. Adam, Y. Qu, J.W. Davis, M.D. Ward, M.A. Clements, L.H. Cazares et al., Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men, *Cancer Res* **62**(13) (2002), 3609–3614.
- [2] American Cancer Society, Cancer Facts and Figures 2003.
- [3] M.C. Beduschi and J.E. Osterling, Percent free prostate-specific antigen: the next frontier in prostate-specific antigen testing, *Urology* **51** (1998), 98–109.
- [4] M.K. Brawer, M.P. Chetner, J. Beatie et al., Screening for prostatic carcinoma with prostate carcinoma with prostate specific antigen, *J Urol* **147** (1992), 841–845.
- [5] P. Bunting, A. Scorilas, L.Y. Luo et al., Human kallikrein 10: a novel tumor marker for ovarian carcinoma? *Clin Chem Acta* **306** (2001), 111–118.
- [6] D.J. Cahill, Protein and antibody arrays and their medical applications, *J Immunol Meth* **250** (2001), 81–91.
- [7] W.J. Catalona, J.P. Richie, F.R. Ahmann et al., Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men, *J Urol* **151** (1994), 1283–1290.
- [8] W.J. Catalona, D.S. Smith, T.L. Ratliff et al., Measurement of prostate-specific antigen in serum as a screening test for prostate cancer, *N Engl J Med* **324** (1991), 1156–1161.
- [9] W.J. Catalona, D.S. Smith, T.L. Ratliff et al., Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening, *J Am Med Assoc* **270** (1993), 948–954.
- [10] C.A. Coltman Jr, I.M. Thompson Jr. and P. Feigl, Prostate cancer prevention trial (PCPT) update, *Eur Urol* **35**(5–6) (1999), 544–547.
- [11] W.H. Cooner, B.R. Mosley, C.L. Rutherford Jr. et al., Prostate cancer detection in a clinical urological practice by ultrasonography, digital rectal examination and prostate specific antigen, *J Urol* **143** (1990), 1146–1152.

- [12] A.V. D'Amico, R. Whittington, S.B. Malkowicz, K. Cote, M. Loffredo, D. Schultz, M.H. Chen, J.E. Tomaszewski, A.A. Renshaw, A. Wein and J.P. Richie, Biochemical outcome after radical prostatectomy or external beam radiation therapy for patients with clinically localized prostate carcinoma in the prostate specific antigen era, *Cancer* **95**(2) (2002), 281–286.
- [13] E. Engvall and P. Perlmann, Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G, *Immunochemistry* **8** (1971), 871–874.
- [14] H. Ge, UPA, a universal protein array system for quantitative detection of protein – protein, protein – DNA, protein – RNA and protein – ligand interactions, *Nuc Acid Res* **28**(2) (2000), i–vii e3.
- [15] B.F. Hankey, E.J. Feuer, L.X. Clegg et al., Cancer surveillance series: interpreting trends in prostate cancer-part I: evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates, *J Natl Cancer Inst* **91** (1999), 1017–1024.
- [16] M. Hara, Y. Koyanagi, T. Inoue et al., Some physico-chemical characteristics of gamma-seminoprotein, an antigenic component specific for human seminal plasma, *Jap J Legal Med* **25** (1971), 322–324.
- [17] F. Labrie, A. Dupont, R. Suburu et al., Serum prostate specific antigen as pre-screening test for prostate cancer, *J Urol* **147** (1992), 846–851.
- [18] A.W. Partin and J.E. Oesterling, The clinical usefulness of prostate specific antigen: update 1994, *J Urol* **152** (1994), 1358–1368.
- [19] C.P. Paweletz, B. Trock, M. Pennanen, T. Tsangaris, C. Mag-nant, L.A. Liotta et al., Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer, *Disease Markers* **17** (2001), 301–307.
- [20] M.S. Pepe, R. Etzioni, Z. Feng, J.D. Potter, M.L. Thompson, M. Thornquist et al., Phases of biomarker development for early detection of cancer, *J Natl Cancer Inst* **93** (2001), 1054–1061.
- [21] M. Perrotti, F. Rabbani, A. Farkas et al., Trends in poorly differentiated prostate cancer 1973 to 1994: observations from the Surveillance, Epidemiology and End Results database, *J Urol* **160** (1998), 811–815.
- [22] E.F. Petricoin III, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg et al., Use of proteomic patterns in serum to identify ovarian cancer, *Lancet* **39** (2002), 572–577.
- [23] Y. Qu, B.-L. Adam, Y. Yasui, M.D. Ward, L.H. Cazares, P.F. Schellhammer et al., Boosted decision tree analysis of SELDI mass spectral serum profiles discriminates prostate cancer from non-cancer patients, *Clin Chem* **48**(10) (2002), 1835–1843.
- [24] G. Ramsay, DNA chips: state-of-the art, *Nat Biotech* **16** (1998), 40–44.
- [25] B. Schwetzer and S. Kingsmore, Measuring proteins on microarrays, *Curr Opin Biotech* **13** (2002), 14–19.
- [26] SEER Cancer Progress Report, 2001; see figure 22; <http://progressreport.cancer.gov/>.
- [27] D. Shalon, S.J. Smith and P.O. Brown, A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization, *Genome Res* **6** (1996), 639–645.
- [28] R.E. Tarone, K.C. Chu and O.W. Brawley, Implications of stage-specific survival rates in assessing recent declines in prostate cancer mortality rates, *Epidemiology* **11** (2000), 167–170.