

A Novel, High-Throughput Workflow for Discovery and Identification of Serum Carrier Protein-Bound Peptide Biomarker Candidates in Ovarian Cancer Samples

MARY F. LOPEZ,^{1*} ALVYDAS MIKULSKIS,¹ SCOTT KUZDZAL,¹ EVA GOLENKO,¹
EMANUEL F. PETRICOIN III,² LANCE A. LIOTTA,² WAYNE F. PATTON,¹ GORDON R. WHITELEY,³
KEVIN ROSENBLATT,⁴ PREM GURNANI,⁴ ANIMESH NANDI,⁴ SAMUEL NEILL,⁵ STUART CULLEN,⁵
MARTIN O'GORMAN,⁵ DAVID SARRACINO,⁶ CHRISTOPHER LYNCH,¹ ANDREW JOHNSON,¹
WILLIAM MCKENZIE,¹ and DAVID FISHMAN⁶

Background: Most cases of ovarian cancer are detected at later stages when the 5-year survival is ~15%, but 5-year survival approaches 90% when the cancer is detected early (stage I). To use mass spectrometry (MS) of serum proteins for early detection, a seamless workflow is needed that provides an opportunity for rapid profiling along with direct identification of the underpinning ions.

Methods: We used carrier protein-bound affinity enrichment of serum samples directly coupled with MALDI orthogonal TOF MS profiling to rapidly search for potential ion signatures that contained discriminatory power. These ions were subsequently directly subjected to tandem MS for sequence identification.

Results: We discovered several biomarker panels that enabled differentiation of stage I ovarian cancer from unaffected (age-matched) patients with no evidence of

ovarian cancer, with positive results in >93% of samples from patients with disease-negative results and in 97% of disease-free controls. The carrier protein-based approach identified additional protein fragments, many from low-abundance proteins or proteins not previously seen in serum.

Conclusions: This workflow system using a highly reproducible, high-resolution MALDI-TOF platform enables rapid enrichment and profiling of large numbers of clinical samples for discovery of ion signatures and integration of direct sequencing and identification of the ions without need for additional offline, time-consuming purification strategies.

© 2007 American Association for Clinical Chemistry

The need for development of early screening methods for epithelial ovarian carcinoma is particularly urgent. Ovarian cancer is the 4th leading cause of cancer-related deaths among women in the United States (1, 2). Regrettably, 70%–75% of new cases are diagnosed in stage III or IV, with a predicted 5-year survival of ~15% (3). The 5-year survival approaches 90%, however, if cancer is detected when confined to the ovary (stage I) (3). More sensitive and specific tests for earlier detection may improve patient survival rates by facilitating early treatments such as surgical intervention (4–8).

Recently, new methods of disease detection based on discriminant mass spectral analysis (serum pattern profiling) have been proposed (9–14). The power of this approach is 4-fold: (a) it is unbiased and does not presuppose any particular disease mechanism, (b) multiple differences, i.e., multiple putative disease markers, are often discovered and combinations of markers are likely to be

¹ PerkinElmer Life and Analytical Sciences, Wellesley, MA.

² Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, VA.

³ Clinical Proteomics Reference Laboratory, Gaithersburg, MD.

⁴ Department of Pathology, Division of Translational Pathology, University of Texas Southwestern Medical Center, Dallas, TX.

⁵ Nonlinear Dynamics, Cuthbert House, All Saints, Newcastle upon Tyne, United Kingdom.

⁶ Harvard Partners, Cambridge, MA.

⁷ New York University Medical School, Division of Obstetrics and Gynecology, New York, NY.

* Address correspondence to this author at: PerkinElmer Life and Analytical Sciences, 45 Williams St., Wellesley, MA 02481-4008. Fax 617-574-9864; e-mail mary.lopez@perkinelmer.com.

Received September 29, 2006; accepted March 27, 2007.

Previously published online at DOI: 10.1373/clinchem.2006.080721

more powerful discriminators than single markers, (c) large numbers of samples (usually blood sera) from appropriate cohorts can be analyzed quickly for discovery and subsequent validation of putative marker sets, and (d) the method does not require a priori antibody development for success.

Serum peptide pattern profiling studies have shown promise, especially in clinical research (13, 14). The approach has, however, generated controversy centered primarily on 2 issues: the relative importance of obtaining definitive sequence identification for differentiating masses and the likelihood that peptides or protein fragments, as opposed to intact proteins, can be useful as disease biomarkers (14–17). The clinical relevance of the serum peptidome has been vigorously debated, but recent publications have confirmed that specific protein fragments are correlated with disease stages (18, 19). A major area of controversy has been the lack of data consistency and reproducibility across the various published studies (17), although with proper attention to stringent experimental design and protocols, these issues can be addressed in future studies (20). High resolution and mass accuracy are required for accurate comparison of mass spectral data across a broad mass range. Sample preparation is another area of critical importance. Reduction of sample complexity is an essential first step for all blood-borne biomarker discovery methods because of the large dynamic range of protein concentrations (21). Low-throughput methods such as 2-dimensional gels coupled with mass spectrometry (MS)⁸ or shotgun proteomics have been used to mine the serum proteome for biomarkers (22–25), but these methods have limited clinical value because they do not allow the simultaneous screening of adequate (large) numbers of samples. Numerous depletion strategies have been developed to remove the more common, abundant proteins in sera or plasma (26). However, many of these proteins (in part because of their high abundance in the blood), act as carrier proteins and bind a vast assortment of peptides and protein fragments (27). With an albumin blood concentration >600 $\mu\text{mol/L}$, the probability is >98% that even molecules with relatively low binding affinities will be complexed with albumin (19). These carrier protein-bound protein fragments and peptides may provide potential diagnostic information for many diseases (18, 28, 29). Peptides and protein fragments from proteins catabolized by proteolytic cascades in tissues diffuse into the blood and are bound to highly abundant proteins such as albumin. This process effectively prolongs the bloodstream half-life of low molecular weight peptides and protein fragments that otherwise would be eliminated by the kidneys. The collection of peptides and protein fragments bound to carrier proteins therefore provides a metabolic snapshot

of diseased and normal tissues, and the low molecular weight peptide archive can be viewed as a direct reflection of the ongoing pathophysiology that can facilitate a true systems biology approach to biomarker discovery. Thus, adsorption to albumin serendipitously provides an endogenous and very efficient enrichment process for rare or low-abundance protein fragments. Unfortunately, many potentially interesting biomarkers are likely to be inadvertently eliminated during commonly used strategies for the removal of albumin and other highly abundant proteins in blood serum, and sample fractionation protocols that omit the capture of carrier protein-bound peptides may yield preparations consisting almost exclusively of peptides derived from blood coagulation cascades (30, 31).

To develop a new workflow for biomarker candidate identification, we evaluated high-throughput carrier protein-bound affinity enrichment of serum samples coupled with high-resolution MALDI orthogonal TOF (OTOF) MS, discriminant analysis of the resulting mass spectral patterns, and sequence identification of the discriminating ions to search for putative early protein/peptide biomarkers in ovarian cancer serum samples.

Materials and Methods

CLINICAL SERUM SAMPLES

Serum samples were obtained from study participants with full consent and institutional review board approval and collected before physical evaluation, diagnosis, and treatment. The study set samples are shown in Table 1. Serum samples from cancer patients were obtained from the National Ovarian Cancer Early Detection Program and the gynecologic oncology clinic at Northwestern University (Chicago, IL). Cancer samples were stage I–IV. The apparently healthy (nondisease) serum samples were collected from Northwestern University; Innsbruck Medical University, Innsbruck, Austria; and Johns Hopkins University, Baltimore, MD, from age-matched women without any evidence of cancer for 5 years before sample collection. All samples were collected from 1999 to 2002 and stored until analysis in 2005. The mean age for all study participants (cancer patients and healthy individuals) was ~40 years for all groups. To minimize institutional bias attributable to sample handling techniques (15), all samples were collected and handled with the same standard operating procedures. Ovarian cancer samples were obtained from patients before therapy and surgery; all patients were surgically staged. There was a predominance of serous cases (54.9%) with 7.1% clear cell

Table 1. Clinical samples.^a

	Healthy	Total cancer	High CA125	Low CA125
Samples	110	453	214	239
Spectra	330	1359	642	717

^a Samples were classified as high CA125 if the concentrations were >22 kIU/L, and low CA125 if <22 kIU/L (see *Materials and Methods*).

⁸ Nonstandard abbreviations: MS, mass spectrometry; OTOF, orthogonal TOF.

and the remaining 38.0% listed only as adenocarcinoma. Each sample was accompanied by a verified pathologic diagnosis.

All serum samples were processed from blood drawn under strict National Cancer Institute/US Food and Drug Administration Proteomics Program standard operating guidelines as follows: Specimens were collected in red-top Vacutainer tubes and allowed to clot for 1 h on ice, followed by centrifugation at 4 °C for 10 min at 2000g. The serum supernatant was divided into aliquots and stored at -80 °C until needed. The serum samples were assayed for CA125 by use of the Elecsys CA125 reagent set (Roche Pharmaceuticals). Samples with a CA125 concentration >22 kIU/L were classified as high CA125 and those with a concentration <22 kIU/L as low CA125. We selected 22 kIU/L because it was the mean of the samples in the study; CA125 >35 kIU/L is generally considered to be the cutoff indicating likely disease recurrence. CA125 was below the limit of detection in the healthy controls.

SAMPLE PROCESSING

Samples from cancer patients and healthy controls were processed in a random order to account for any systematic errors and variations from experiment to experiment. Locations of all samples were random on each MALDI plate.

Serum samples were processed using prototype ProXPRESSION[™] biomarker enrichment reagent sets (PerkinElmer) as described in (28). The Cibachron blue dye affinity chromatography-based technology is designed to capture high-abundance carrier proteins in blood (such as albumin) and dramatically enriches the peptide and protein fragments bound to the carrier proteins. ZipPlates[™] and a vacuum manifold were purchased from Millipore. Millipore also provided custom-fitting adapters for direct spotting of samples on single use MALDIchip[™] prOTOF Target plates (PerkinElmer). Premixed alpha-cyano-4-hydroxycinnamic acid matrix was from PerkinElmer.

MALDI-OTOF MS

Mass spectra were acquired on a prOTOF[™] 2000 MALDI-OTOF Mass Spectrometer interfaced with TOFWorks[™] software (PerkinElmer/SCIEX). Because of the orthogonal design, a single external mass calibrant was used to achieve better than 5 ppm mass accuracy over an entire sample plate (up to 384 samples). In this study, a 2-point external calibration of the prOTOF instrument was performed before acquiring the spectra in a batch mode from 96 samples. MALDI-OTOF-MS can collect data over a wide range of mass values (300 kDa) in a single acquisition. Typical resolution for peptides and proteins up to 10 kDa was >12 000 full width at half maximum.

The raw mass spectral data used in this study are accessible without restriction upon request to the corresponding authors.

ULTRAHIGH RESOLUTION TANDEM MS

Pools (6) of either ovarian cancer samples or healthy samples were dissolved in 50 μ L of 5% acetonitrile 0.1% formic acid/water, and transferred to an MS plate and lyophilized. Samples in 5% acetonitrile 0.1% formic acid were injected with a Famos Autosampler onto a 75 μ m \times 18 cm fused silica capillary column packed with C18 or C8 media, in a 250- μ L/min gradient of 5% acetonitrile 0.1% formic acid to 50% acetonitrile 0.1% formic acid over the course of 100 min with a total run length of 150 min. For the 240-min runs, 25-cm columns were used to achieve higher chromatographic resolution and loading capacity. The LTQ-FT ultrahybrid mass spectrometer (Thermo Fisher Scientific) was run in a top 4 configuration at 200 K resolution for a full scan. Ions that were +1 or undefined in charge states were rejected for MS 2 analysis. Dynamic exclusion was set to 1 with a limit of 180 s, with early expiration set to 6 full scans. Peptide identification was performed with Sequest through the Bioworks Browser 3.2 EF2 (Thermo Scientific). Database searches were made with a no-enzyme indexed version of the National Center for Biotechnology Information RefSeqhuman/reversed Refseqhuman database using differential oxidized methionines at a tolerance of 10 ppm. Peptide score cutoff values were chosen at Xcorr of 1.8 for singly charged ions, 2.0 for doubly charged ions, and 2.5 for triply charged ions, along with deltaCN values of ≥ 0.1 , and rank score preliminary values of <10 with a peptide *P* value of $1e-3$ or better. The small mass tolerance of the search ensured that only relevant peptides were matched. The cross-correlation values chosen for each peptide assured a high confidence match for the different charge states, and the deltaCN cutoff insured the uniqueness of the peptide hit. The *P* value is a probability score for a random hit peptide. Typically, multiple peptide hits were obtained for any identified protein, for example, more than 28 separate hits were obtained for plasma kallikrein-sensitive glycoprotein (data not shown). However, there were also a number of proteins that were identified by single peptide hits.

PROCESSING AND ANALYSIS OF SPECTRAL PROFILES

Progenesis PG600[™] software (NonLinear Dynamics) was used to process and analyze the OTOF mass spectral data. Raw spectra from the OTOF were directly loaded into the PG600 program using the prOTOF loader program. Binning was set at 4. Analyses were performed to find discriminant markers between the following groups: healthy vs all cancer, healthy vs high CA125, healthy vs low CA125, and healthy vs stage I cancer. For the initial analysis, the stringency parameters for biomarker selection were set to include peaks with an mean quantity threshold of ≥ 75 (higher intensity peaks, to facilitate subsequent sequence identification by tandem MS) and $P \leq 0.01$. Subsequent analyses were performed with peak intensity stringency of <50 or 0 and a $P \leq 0.01$. These parameters ensured the detection of differently expressed

peaks that were highly significant. Once the putative peaks were detected, classification models were developed using flexible discriminant analysis and the R statistical package (32). The flexible discriminant analysis algorithm determined nonlinear decision boundaries that were better able to separate classes, resulting in a classification technique that was more powerful for high-dimensional data with complex interrelationships. We used independent stratified balanced random sampling to split the data into a training set and a test set. The training set was used to build a classifier model, and this model was evaluated on the test set. The classifier classified test cases as healthy or diseased, and these data were then used to create ROC curves. Monte Carlo cross-validation of training and test sets was used (33,34). For this validation, the results of 100 runs of the sampling and classifier modeling procedure were averaged together to create the final ROC curve.

Results

DEVELOPMENT OF STUDY DESIGN AND WORKFLOW

When processed with the carrier protein-based biomarker enrichment protocol, serum samples routinely generated highly reproducible peptide profiles with intensity CVs of 5%–10%, on average (28). The workflow and study design

are shown in Fig. 1. All samples were processed in a high-throughput, parallel manner to obtain the information-rich mass spectra. Spectra from the various groups were compared and analyzed, and the discriminant peptide masses were identified. Subsequently, pooled disease and healthy serum samples were processed for peptide enrichment and then submitted for de novo sequence analysis by ultrahigh resolution tandem-MS. This procedure was efficient and allowed rapid (hours to days) discovery and identification of putative disease markers from large numbers of samples. Two advantages to this approach were as follows: (i) a large number of samples could be analyzed simultaneously, lending statistical relevance to the putative markers, and (ii) sequence identification was directly obtained from the same samples, increasing confidence in marker identification accuracy and dispelling the need for further purification by gels or other methods.

ANALYSIS AND IDENTIFICATION OF DISCRIMINANT PEPTIDE MASSES AND PUTATIVE BIOMARKERS

An initial set of 9 discriminating peptides resulted from the initial analysis comparing the spectral profiles from 4 different groups: healthy vs all cancer (low CA125 + high CA125), healthy vs low CA125, healthy vs high CA125,

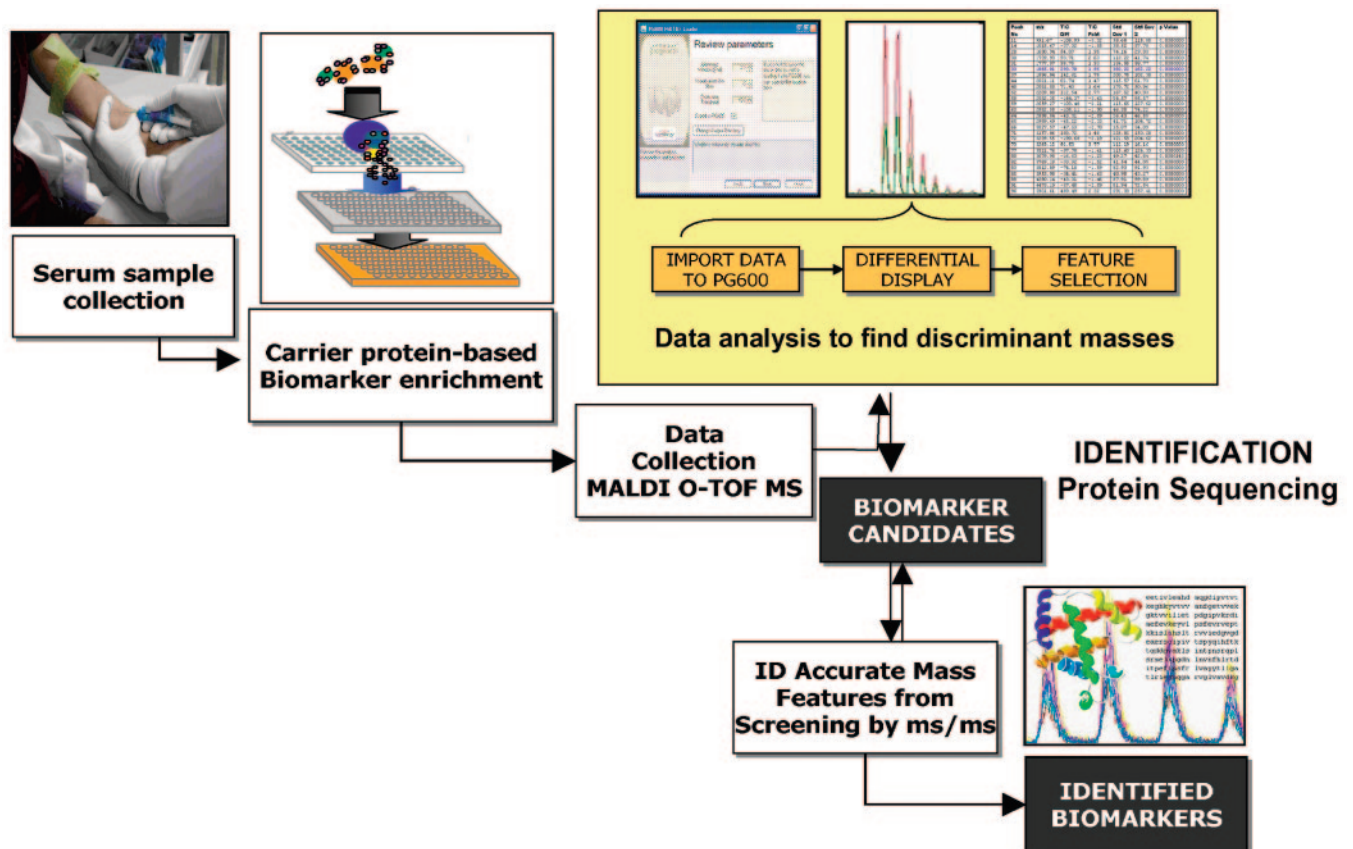


Fig. 1. Serum based biomarker discovery workflow.

Serum is processed through albumin capture ProXPRESSION 96-well plates. Fragments bound to albumin are eluted, concentrated through ZipPlate, and analyzed by MALDI OTOF MS and discriminant analysis. The differentially expressed fragments are identified by sequence analysis using tandem MS.

and healthy vs stage I cancer (Table 2). The identities of these peptides included multiple peptide hits from complement component 3 and interalpha (globulin) inhibitor H4, and single peptides from complement component 4A, transthyretin, and fibrinogen (Table 3). Because the analysis parameters were purposefully set to screen out low-intensity masses (to improve our success rate with sequence identification), it is not surprising that these fragments are derived from relatively abundant serum proteins. Although these peptides were derived from highly abundant resident proteins, they are not necessarily nonspecific or themselves highly abundant. Examples of overlaid disease and healthy spectra are shown in Fig. 2; putative marker expression ranged from an ~3.6-fold increase to a 2.6-fold decrease in cancer samples (Table 2). To test the discriminating power of the 9-marker set, a flexible discriminant analysis classification model was built and tested. The discriminating power of the 9-marker model was quite good; 93% of samples from patients with disease were positive and 93% of samples from disease-free controls were negative (Table 2).

To expand the analysis, we reduced the stringency parameter slightly to allow the inclusion of lower intensity discriminating masses. The results of this analysis yielded the set of 4 markers shown in Tables 2 and 3. Two of the peptides at m/z 1739.9 and 2582.35 were also in the initial set of 9, and the remaining 2 masses at m/z 2659.27 and 2989.49 remained unidentified. The 4-marker model delivered equivalent diagnostic sensitivity (93%) and better specificity (97%) than the 9-marker model (Table 2). In an effort to discover lower abundance discriminating

peptides, or peptides not related to coagulation, we lowered the intensity stringency of the analysis to 0, keeping the P values at 0.01 or better. Five additional discriminating masses resulted from this analysis (Tables 2 and 3). None of the identified proteins in this set are related to coagulation, and all are correlated with or involved in cellular oncogenesis [casein kinase 2, transgelin (35–37)], proliferation [keratin 2, LARGE (38, 39)], or detoxification of ROS [diamine oxidase (40)]. A model created with these 5 markers plus 2 additional markers from the 4-marker set classified the healthy and low CA125 samples with 77% sensitivity and 85% specificity (Table 2, 7-marker model).

ADDITIONAL PROTEIN FRAGMENTS DETECTED IN CANCER SERA

In addition to the discriminating protein fragments identified above, additional sequence identities were obtained (162 total) for other peptides/proteins in the ovarian cancer serum samples (see Table 1 in the Data Supplement that accompanies the online version of this article at <http://www.clinchem.org/content/vol53/issue6>). Many of these fragments were not detected in the healthy sera. We could not be certain that the fragments were exclusive to the cancer sera, however, because we extensively sequenced peptides from only a limited number of pooled samples. Further sequencing experiments may determine the exclusivity of these peptides to disease or healthy samples and their potential as putative disease biomarkers. Many of the proteins from which these peptides are

Table 2. Healthy vs cancer.^a

Model	m/z	Fold expression	P	Diagnostic sensitivity, % (SD)	Diagnostic specificity, % (SD)	Area under ROC curve	% Prediction error
9	1690.94	-2.06	0.0000000	93 (2.5)	93 (1.7)	0.98 (0.01)	7.0 (1.3)
	1739.93	-2.62	0.0000000				
	1777.97	-2.03	0.0000000				
	1865.01	-2.03	0.0000000				
	2021.11	-1.51	0.0000000				
	2582.35	3.38	0.0000000				
	2898.54	2.01	0.0000000				
	3027.57	2.59	0.0000000				
	3239.55	2.10	0.0000000				
4	1739.93	-2.62	0.0000000	93 (2.0)	97 (0.8)	0.98 (0.01)	4.3 (0.7)
	2582.35	3.38	0.0000000				
	2659.27	2.00	0.0000000				
	2989.49	2.23	0.0000000				
7	1739.93	-2.62	0.0000000	77 (4.3)	85 (3.0)	0.87 (0.02)	16.9 (1.9)
	2582.35	3.38	0.0000000				
	1966.91	-1.21	0.0000000				
	1041.68	1.34	0.0000000				
	2115.05	-1.11	0.0137930				
	1224.68	1.87	0.0000000				
	2345.19	2.07	0.0000000				

^a Marker model expression and classification (healthy vs low CA125 samples).

Table 3. Sequence identifications of putative ovarian cancer peptide biomarkers.

<i>m/z</i>	Peptide sequence	<i>P</i> (pep)	Identity
1041.68	L.NVKVDPEIQ.N	4.1876E-05	gi 47132620 ref NP_000414.2 keratin 2a [Homo sapiens] [MASS = 65432]
1224.68	L.KPRVSWIPNK.H	2.54E-04	gi 33285008 ref NP_689525.2 glycosyltransferase-like 1B [Homo sapiens]
1690.94	S.KITHRIHWESASLL.R	5.40E-10	gi 4557385 ref NP_000055.1 complement component 3 precursor [Homo sapiens]
1739.93	R.NGFKSHALQLNLRQI.R	9.96E-09	gi 67190748 ref NP_009224.2 complement component 4A preproprotein [Homo sapiens]
1777.97	S.SKITHRIHWESASLL.R	3.07E-11	gi 4557385 ref NP_000055.1 complement component 3 precursor [Homo sapiens]
1865.01	R.SSKITHRIHWESASLL.R	1.35E-09	gi 4557385 ref NP_000055.1 complement component 3 precursor [Homo sapiens]
1966.91	T.DVNTHRPREYWDYES.H	3.0179E-05	gi 29570791 ref NP_808227.1 casein kinase II alpha 1 subunit isoform a [Homo sapiens]
2021.11	R.SSKITHRIHWESASLLR.S	1.13E-10	gi 4557385 ref NP_000055.1 complement component 3 precursor [Homo sapiens]
2115.05	A.REGADVIVNCTGVWAGALQR.D	0.00061097	gi 21536470 ref NP_001908.2 D-amino-acid oxidase [Homo sapiens] [MASS = 39496]
2345.19	Q.M*GTNRGASQAGM*TYGGM*PRQIL.-	1.9783E-05	gi 4507357 ref NP_003555.1 transgelin 2 [Homo sapiens] [MASS = 22391]
2582.35	R.NVHSGSTFFKYLLQGAKIPKPEA.S	3.9553E-06	gi 31542984 ref NP_002209.2 inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glyco
2898.54	G.PRRYIAALLSPYSYSTTAVWTPKE.-	0.00173243	gi 4507725 ref NP_000362.1 transthyretin [Homo sapiens] [MASS = 15887]
3027.57	N.FRPGVLSRQLGLPGPPDVPDHAAYHPF.R	2.9701E-06	gi 31542984 ref NP_002209.2 inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glyco
3239.55	K.SYKMADEAGSEADHEGTHSTKRGHAKSRPV.R	0.00140644	gi 11761629 ref NP_068657.1 fibrinogen, alpha chain isoform alpha preproprotein [Homo sapiens]

derived are present in very low abundance or have not been identified previously in blood (41).

Discussion

The primary aim of this study was to investigate the utility of a new workflow method for biomarker discovery. As a means to this end, we used a large and well-characterized set of serum samples that reflected an urgent clinical need. The new workflow we describe for discovery of potentially clinically-relevant biomarkers from large, statistically powered study sets incorporates MS profiling with an extremely stable MALDI-TOF platform coupled to an automated process for albumin-bound peptide enrichment and direct sequencing and identification of ion peaks by high-resolution tandem MS without the need for additional chromatography or enrichment. Thus the discriminating ions are directly identified and sequenced.

The carrier protein-based approach yielded a number of discriminating peptides, and we built flexible discriminant models with multiple marker sets (9, 4, and 7 markers). These models enabled classification with high specificity and sensitivity of samples from cancer patients and healthy controls. Perhaps not surprisingly, peptide fragments associated with the coagulation cascade provided the highest classification power. Substantial evi-

dence supports the association between activation of blood coagulation and progression of cancer. Recent studies from several laboratories have linked malignant transformation (oncogenesis), tumor angiogenesis, and metastasis to the generation of clotting intermediates, clotting or platelet function inhibitors, or fibrinolysis inhibitors (42). Many researchers have published putative markers for cancer, and ovarian cancer in particular, that are related to coagulation or inflammation (43). Interestingly, when we lowered the stringency of our analysis to allow the inclusion of very low intensity discriminating signals, we identified a further 5 peptides not related to either coagulation or inflammation pathways. Among these lower intensity peptides, casein kinase 2 is oncogenic and up-regulated in tumors (35), and transgelin has been reported previously as a putative marker for ovarian and endometrial cancer in other studies using widely different discovery methods including LC-MS and cDNA-representational difference analysis (36, 37). The other 3, keratin 2, glycosyl transferase (LARGE), and diamino oxidase are also associated with processes related to cancer (38–40).

In addition to the discriminating peptides described above, a rich trove of protein fragments, many from low-abundance proteins or proteins not previously seen in serum (41), were recovered from the ovarian cancer sera. These results are consistent with those reported by

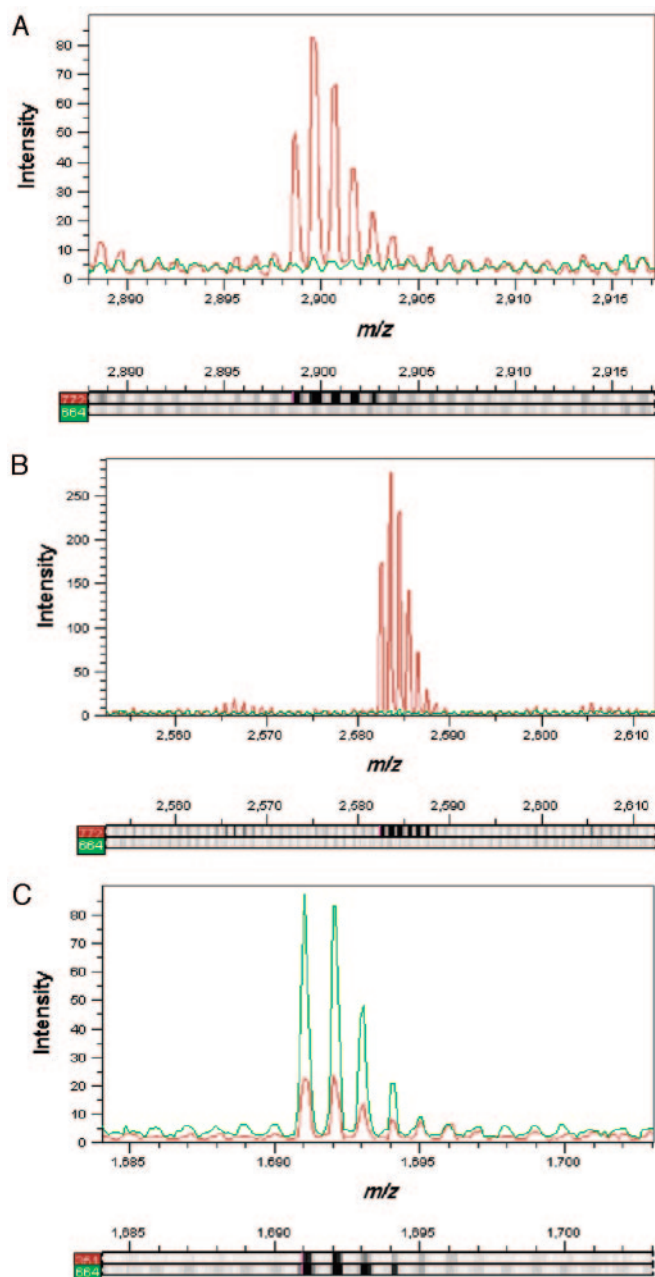


Fig. 2. Differential expression of fragments.

(A), transthyretin fragment at m/z 2898.54, zoom of m/z 2890 to 2915 area. (B), plasma kallikrein-sensitive glycoprotein fragment at m/z 2582.35, zoom of m/z 2550 to 2610 area. (C), complement component precursor 3 fragment at m/z 1690.94, zoom of m/z 1685 to 1700 area.

Screen capture from PG600 analysis. Overlaid mass spectral traces from representative ovarian cancer and healthy samples. Red = disease, green = healthy.

Lowenthal et al. (19). In this study, we identified a number of proteins associated with cellular proliferation, cancer, and cancer signaling pathways in the ovarian cancer samples. Although we could not be certain that these peptides/proteins were exclusive to the cancer sera, many of the peptides were not found in the pooled healthy serum samples. A sampling of these proteins and

their functions are described in Table 2 of the online Data Supplement.

Several proteins identified in this study (e.g. transthyretin) have also been reported by others as putative biomarkers for ovarian cancer (10, 45). Interestingly, the transthyretin fragment reported herein is a different unique mass than that reported in a previous serum-based ovarian cancer study (10).

The proteins we identified are involved in cellular inflammation, differentiation, signaling, apoptosis, transcriptional regulation, and other regulatory mechanisms. It is remarkable that this rich variety of low-abundance species is so well represented in the fraction bound to serum albumin.

In summary, in a period of 2–3 weeks we identified ~162 proteins from peptides and protein fragments bound to carrier proteins from ovarian cancer patient serum samples. Within this study, 3 sets of the discriminating carrier-protein bound fragments differentiated samples from patients with ovarian cancer and from apparently healthy controls with sensitivities and specificities of up to 93% and 97%, respectively. These values compare very favorably with published mean sensitivities and specificities of ~50% for CA125, the current gold standard biomarker for ovarian cancer (4). Thus, this new high-throughput, top-down approach to biomarker discovery provides a clear path for the rapid detection of potential markers for early disease detection.

Grant/funding support: None declared.

Financial disclosures: M.F.L., A.M., S.K., E.G., W.F.P., C.L., A.J., and W.M. are employees of PerkinElmer Life and Analytical Sciences, the vendor of ProXPRESSIONS reagent sets and the pTOF 2000 orthogonal MALDI-TOF mass spectrometer described in this manuscript. L.A.L. and E.F.P. have US government- and University-assigned patents that cover certain aspects of the technology discussed. No other financial interests declared.

References

1. Bast RC. Status of tumor markers in ovarian cancer screening. *J Clin Oncol* 2003;21:200–5.
2. Boyce EA, Kohn EC. Ovarian cancer in the proteomics era: diagnosis, prognosis, and therapeutics targets. *Int J Gynecol Cancer* 2005;15(Suppl 3):266–73.
3. Bhoola S, Hoskins WJ. Diagnosis and management of epithelial ovarian cancer. *Obstet Gynecol* 2006;107:1399–410.
4. Moss EL, Hollingworth J, Reynolds TM. The role of CA125 in clinical practice. *J Clin Pathol* 2005;58:308–12.
5. Hortin GL, Jortani SA, Ritchee, JC Jr, Valdes R Jr, Chan DW. Proteomics: a new diagnostic frontier. *Clin Chem* 2006;52:1218–22.
6. Wulfkuhle JD, Liotta LA, Petricoin EF. Proteomic applications for the early detection of cancer. *Nat Rev Cancer* 2003;3:267–75.
7. Ardekani AM, Liotta LA, Petricoin EF III. Clinical potential of proteomics in the diagnosis of ovarian cancer. *Expert Rev Mol Diagn* 2002;2:312–20.

8. Srinivas PR, S. Srivastava S, Hanash S, Wright GL. Proteomics in early detection of cancer. *Clin Chem* 2001;47:1901–11.
9. Rosenblatt KP, Bryant-Greenwood P, Killian JK, Mehta A, Geho D, Espina V, et al. Serum proteomics in cancer diagnosis and management. *Annu Rev Med* 2004;55:97–112.
10. Zhang Z, Bast RC Jr, Yu Y, Li J, Sokoll LJ, Rai AJ, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res* 2004;64:5882–90.
11. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan DW. Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* 2002;48:1296–1304.
12. Petricoin EF, Ardekani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002;359:572–7.
13. Hortin GL. The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin Chem* 2006;52:1223–37.
14. Whiteley GR. Proteomic patterns for cancer diagnosis-promise and challenges. *Mol BioSyst* 2006;2:358–63.
15. Baggerly KA, Morris JS, Edmonson SR, Coombes KR. Signal in noise: evaluating reported reproducibility of serum proteomic tests for ovarian cancer. *J Natl Cancer Inst* 2005;97:307–9.
16. Liotta LA, Lowenthal M, Mehta A, Conrads TP, Veenstra TD, Fishman DA, et al. Importance of communication between producers and consumers of publicly available experimental data. *J Natl Cancer Inst* 2005;97:310–14.
17. Diamandis EP. Point: proteomic patterns in biological fluids: do they represent the future of cancer diagnostics? *Clin Chem* 2003;49:1272–5.
18. Liotta LA, Petricoin EF. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. *J Clin Invest* 2006;116:26–30.
19. Lowenthal MS, Mehta AI, Frogale K, Bandle RW, Araujo RP, Hood BL, et al. Analysis of albumin-associated peptides and proteins from ovarian cancer patients. *Clin Chem* 2005;51:1933–45.
20. Diamandis EP. Peptidomics for cancer diagnosis: present and future. *J Proteome Res* 2006;9:2079–82.
21. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the low molecular weight human serum proteome. *Mol Cell Proteomics* 2003;2:1096–103.
22. Liu H, Sadygov RG, Yates JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004;76:4193–201.
23. Pieper R, Gatlin CL, Makusky AJ, Russo PS, Schatz CR, Miller SS, et al. The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* 2003;3:1345–64.
24. Adkins JN, Varnum SM, Auberry KJ, Moore RJ, Angell NH, Smith RD, et al. Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol Cell Proteomics* 2002;1:947–55.
25. Washburn MP, Wolters D, Yates JR III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19:242–7.
26. Wang YY, Cheng P, Chan DW. A simple affinity spin tube filter method for removing high-abundant common proteins or enriching low-abundant biomarkers for serum proteomic analysis. *Proteomics* 2003;3:243–8.
27. Mehta AI, Ross S, Lowenthal MS, Fusaro V, Fishman DA, Petricoin EF III, et al. Biomarker amplification by serum carrier protein binding. *Dis Markers* 2003;19:1–10.
28. Lopez MF, Mikulskis A, Kuzdzal S, Bennett D, Kelly J, Golenko E, et al. High-resolution serum proteomic profiling of Alzheimer's disease samples reveals disease-specific, carrier-protein-bound mass signatures. *Clin Chem* 2005;51:1946–54.
29. Zhou M, Lucas DA, Chan KC, Issaq HJ, Petricoin EF III, Liotta LA, et al. An investigation into the human serum "interactome". *Electrophoresis* 2004;25:1289–98.
30. Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* 2006;116:271–84.
31. Koomen JM, Li D, Xiao LC, Liu TC, Coombes KR, Abbruzzese J, et al. Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. *J Proteome Res* 2005;4:972–81.
32. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCr: An R Package for visualizing the performance of scoring classifiers. <http://rocr.bioinf.mpi-sb.mpg.de> (accessed January 2006).
33. Hastie TJ, Tibshirani R, Buja A. Flexible Discriminant Analysis by Optimal Scoring. *J Am Stat Assoc* 1994;89:1255–70.
34. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;21:3294–3300.
35. Scaglioni PP, Yung TM, Cai LF, Erdjument-Bromage H, Kaufman AJ, Singh B, et al. A CK2-dependent mechanism for degradation of the PML tumor suppressor. *Cell* 2006;126:269–83.
36. DeSouza L, Diehl G, Rodrigues MJ, Guo J, Romaschin AD, Colgan TJ, et al. Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J Proteome Res* 2005;4:377–86.
37. Chen H, Wang M, Wang XY, Gao S, Wang J, Guan XM. Identification of differential genes in ovarian cancer using representational difference analysis of cDNA. *Chin Med Sci J* 2005;3:185–9.
38. Sieja K, Stanosz S, von Mach-Szczypinski J, Olewniczak S, Stanosz M. Concentration of histamine in serum and tissues of the primary ductal breast cancers in women. *Breast* 2005;14:236–41.
39. Bloor BK, Tidman N, Leigh IM, Odell E, Dogan B, Wollina U, et al. Expression of keratin K2e in cutaneous and oral lesions: association with keratinocyte activation, proliferation, and keratinization. *Am J Pathol* 2003;162:963–75.
40. Juang CM, Yang YH, Lo WH, Lai CR, Hsieh SL, Yuan CC. Altered mRNA expressions of sialyltransferases in ovarian cancers. *Wang PH, Lee WL, Gynecol Oncol* 2005;99:631–9.
41. Anderson NL, Polanski M, Pieper R, Gatlin T, Tirumalai RS, Conrads TP, et al. The human plasma proteome. *Molec Cell Proteom* 2004;3:311–26.
42. Rickles FR. Mechanisms of cancer-induced thrombosis in cancer. *Pathophysiol Haemost Thromb* 2006;35:103–10.
43. Bast RC Jr, Badgwell D, Lu Z, Marquez R, Rosen D, Liu J, et al. New tumor markers: CA125 and beyond. *Int J Gynecol Cancer*. 2005;15(Suppl 3):274–81.
44. Kumar R, Hung M-C. Signaling intricacies take center stage in cancer cells. *Cancer Res* 2005;65:2511–5.
45. Kozak KR, Su F, Whitelegge JP, Faull K, Reddy S, Farias-Eisner R. Characterization of serum biomarkers for detection of early stage ovarian cancer. *Proteomics* 2005;17:4589–96.