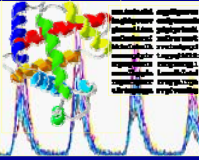


Bridging the Gap between High-Throughput MALDI Screening & Accurate MS/MS Sequence Identifications

M. Lopez, S. Kuzdzal, A. Johnson, W. Patton (PerkinElmer, MA), D. Sarracino (Harvard-Partners Center for Genetics & Genomics, MA)

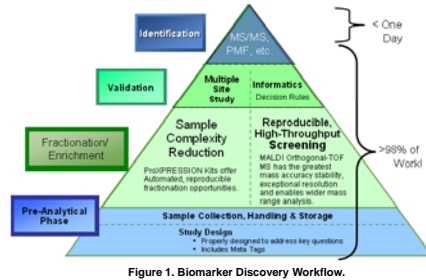


OVERVIEW

Matrix-Assisted Laser Desorption/Ionization (MALDI) mass spectrometry provides researchers the ability to rapidly screen thousands of samples and quickly discern differentially-expressed biomarker candidates. While MALDI MS remains a driving force in biomarker discovery, making the transition from interesting 'profiles' to sequence identifications is typically challenging. Recent advances in sample complexity reduction, MALDI instrument performance and bioinformatics software have enabled researchers to bridge the screening and identification gap. The combination of scalable, membrane-based fractionation strategies (Vivacience/Sartorius AG), enhanced mass accuracy stability and resolution using MALDI Orthogonal-TOF MS (PerkinElmer SCIEX) and new bioinformatics tools (Nonlinear Dynamics, UK) has resulted in the ability to scale-up interesting biomarker differences discovered from high-throughput MALDI O-TOF MS screening to facilitate sequence identifications by MS/MS. This process results in accurate sequence information for differentially-expressed, accurate-mass protein/peptide biomarkers. Examples of FT-ICR MS sequence identifications obtained from ovarian cancer serum biomarker MALDI screening studies will be presented.

INTRODUCTION

In an effort to develop a new "top down" workflow for biomarker candidate identification, we evaluated high-throughput carrier protein-bound affinity enrichment of serum samples coupled with high resolution MALDI orthogonal time of flight (OTOF) mass spectrometry, discriminate analysis of the resulting mass spectral patterns and sequence identification of the discriminating ions to search for putative early protein/peptide biomarkers in serum samples from ovarian cancer patients. Advantages of this discovery platform include scalable, membrane-based fractionation strategies, enhanced mass accuracy stability and new mass spectral pattern analysis tools facilitating the rapid comparison of hundreds or thousands of samples in a highly reproducible and automated manner. Once the relevant differences are identified, the peptides are sequence identified by tandem mass spectrometry facilitating the future development of antibodies and antibody-based immunoassays, including multiplexed immuno-MS for validation and potential diagnostic use. This workflow provides, for the first time, a seamless approach that combines the attractive attributes of an MS based method (speed, label-free analysis, and the ability to analyze without the a priori development of an antibody) with the direct amino acid sequencing and identification of the ion peaks that comprise the discriminatory MS fingerprint.



METHODS

The workflow and study design are shown in Figure 1. Samples were processed in a high-throughput, parallel manner to obtain the information-rich mass spectra (Figure 2). 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 ultra high 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. Advantages to this approach are: a) large numbers of samples can be analyzed simultaneously lending statistical relevance to the putative markers, b) sequence identification is directly obtained from the same samples increasing confidence in marker ID accuracy and dispelling the need for further purification by gels or other methods.

Serum samples were obtained with full subject consent and IRB approval and collected prior to physical evaluation, diagnosis and treatment. Cancer and control samples were processed in a random order to account for any systematic errors and variations from experiment to experiment. Serum samples were processed using prototype ProXPRESSION™ biomarker enrichment kits (PerkinElmer, Boston, MA) (3). This Cibacron blue (CB) dye affinity chromatography-based technology is designed to capture high-abundance carrier proteins in blood (such as albumin) and dramatically enriches for the peptide and protein fragments bound to the carrier proteins. Mass spectra were acquired on a PROTOF™ 2000 MALDI O-TOF Mass Spectrometer (PerkinElmer/SCIEX, Concord, ON, Canada). Due to 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). Progenesis PG600 software (NonLinear Dynamics, Newcastle, UK) was used to process and analyze the OTOF mass spectral data to find statistically significant discriminating peptides. Peptides were sequence identified by LC MS/MS. Ultra-high-resolution tandem mass spectrometry was carried out on an LTQ-FT (ThermoElectron, Waltham, MA) run in a top 4 configuration at 200K resolution for a full scan.

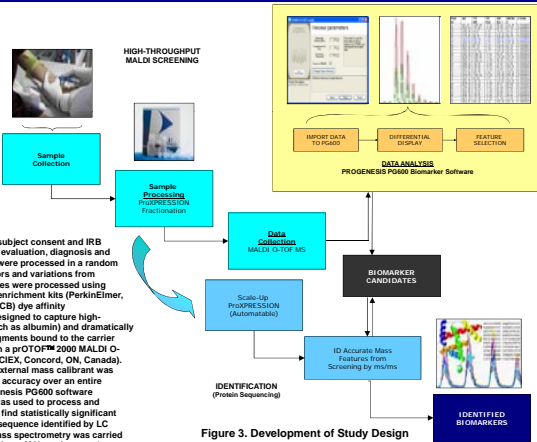
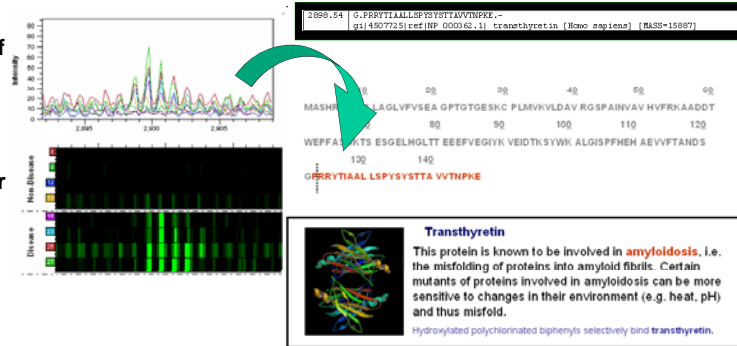


Figure 3. Development of Study Design and Workflow

RESULTS

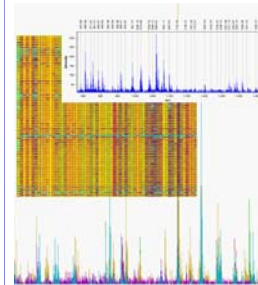
The set of discriminating peptides resulting from the initial analysis of spectral profiles from healthy vs stage I cancer is shown in Table 1. The identities of these peptides included multiple peptide hits from complement component 3 and inter alpha (globulin) inhibitor H4 and single peptides from complement component 4A, transthyretin and fibrinogen (Table 1).



m/z	peptide sequence	P (pep)	identity
1966.91	T.DVNTHRPREYWDYES.H	3.01794E-05	gi 29570791 ref NP_808227.1 casein kinase II alpha 1 subunit isoform a [Homo sapiens]
1041.68	L.NVIVDPEIQ.N	4.18765E-05	gi 47132620 ref NP_000414.2 keratin 2a [Homo sapiens] [MASS=65432]
2115.05	A.REGADVIVNCTGVWAGALQR.D	0.000610972	gi 21536470 ref NP_001908.2 D-amino-acid oxidase [Homo sapiens] [MASS=39496]
1224.68	L.KPRVSWIPNK.H	2.54E-04	gi 33285008 ref NP_689525.2 glycosyltransferase-like 1B [Homo sapiens] [MASS=2345.19]
2345.19	Q.NFGTRNSAQAGMTCVGMPRQL.-	1.97832E-05	gi 4507357 ref NP_003555.1 transgelin 2 [Homo sapiens] [MASS=22391]
1739.93	R.NGFKSHALQLNRLR	9.96E-09	gi 67190748 ref NP_009224.2 complement component 4A preproprotein [Homo sapiens]
1690.84	S.KITHRIHWESALLR	5.40E-10	gi 4557385 ref NP_000055.1 complement component 3 precursor [Homo sapiens] [MA
1865.01	R.SSKTHRIHWESALLR	1.35E-09	gi 4557385 ref NP_000055.1 complement component 3 precursor [Homo sapiens] [MA
2021.11	R.SSKTHRIHWESALLR.S	1.13E-10	gi 4557385 ref NP_000055.1 complement component 3 precursor [Homo sapiens] [MA
1777.97	S.KITHRIHWESALLR	3.07E-11	gi 4557385 ref NP_000055.1 complement component 3 precursor [Homo sapiens] [MA
3027.57	N.FRPGVLSRQLGLRQDPMDHAAHYHFF.R	2.97007E-06	gi 31542984 ref NP_002209.2 inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glyco
2582.35	R.NVHSGSTFFKYVLCAGKMKPEA.S	3.9553E-06	gi 31542984 ref NP_002209.2 inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glyco
3239.55	K.SYKMADEAGSEADHGHTSKRGHAKSRP.VR	0.001406444	gi 11761629 ref NP_068657.1 fibrinogen, alpha chain isoform alpha preproprotein [Homo sapiens] [MA
2898.54	G.PRRYTIAALLSPYSYSTAVVTNPK.E.-	0.001732434	gi 4507725 ref NP_000362.1 transthyretin [Homo sapiens] [MASS=15887]

Table 1. Sequence ID's of putative ovarian cancer peptide biomarkers.

CONCLUSIONS



- Spectral profiles generated with the BioXPRESSION Platform (PerkinElmer) enable rapid biomarker discovery from large clinical sample sets
- Analysis of carrier protein-bound fractions yields potentially valuable diagnostic information
- Putative biomarker protein fragments can subsequently be identified by *de novo* sequencing
- Isobaric Mass Tags provide relative quantification
- After validation, biomarker candidates may be validated and ported to immunoassay (single plex) and immuno ms (multiplex) for the potential development of diagnostic tests