Mass Spectrometry in Cancer Diagnostics

Gunjan Malik, Ph.D.

Postdoctoral Fellow
The George L. Wright Jr. Center for Biomedical Proteomics,
Department of Microbiology and Molecular Cell Biology

Eastern Virginia Medical School Norfolk, VA
The mission of the Center is to discover a means to detect and accurately diagnose a variety of cancers long before the disease becomes life-threatening. Using mass spectrometry-based techniques to visualize and identify proteins, we hope to discover biomarkers that will detect cancer at its earliest stage, as well as markers that can predict disease and treatment outcome.
Some Facts: Why proteomics?

• More than 98% of all diseases are caused by multiple molecular alterations.
• There is very low correlation between mRNA abundance and protein level. (e.g., yeast – *S. cerevisiae*: <0.5 correlation factor)
• Gene products are modified by
  – complex gene interactions
  – cellular events
  – environmental influences
  (co- & post-translational modifications: >189 types)
Pieces of the Puzzle

- Data acquisition, pre-processing, processing, signal versus noise
- Data interpretation phase 1, which signals are potential biomarkers
- Determining sources of variability, response to bias, analytical reproducibility
- Data interpretation phase 2, biomarker maturation
CIPHERGEN PROTEINCHIP® TECHNOLOGY

SELDI ProteinChip® analysis is a mass spectrometric technology that accomplishes protein separation on a chromatographic chip surface by binding subsets of proteins in complex mixtures such as serum. Differentially expressed proteins are determined by comparing protein peak intensity within mass spectra.
The ProteinChip discovery platform

Rapid discovery of biomarkers:
- Serum
- Urine
- Cell lysates
- Cells
- Cerebrospinal fluid
- Tissue homogenates

Purification:
- On-chip
- Micro-chromatography spin columns
- Via traditional purification strategies

Characterization and identification through:
- Epitope mapping
- Peptide mapping
- Phosphorylation
- Peptide sequencing
- Glycosylation analysis
- Binding domain analysis

ProteinChip array is the assay.
- Antibody ProteinChip for low CV’s and accurate quantity
Samples and Requirements:

**Body Fluids:**
- Serum - 20µl
- Urine - 60µl
- Seminal Plasma - 20µl
- Cerebral Spinal Fluid - 60µl

**Cell and Tissue Lysates:**
- Microdissected cells - 2000-3000
- Tissue Culture cells - whole cell lysates
- Subcellular fractionation,
  (i.e. Mitochondrial or membrane lysates)
PixCell™
Laser Capture
Microdissection
Microscope

Before LCM
After LCM
Captured Cells

Protein Extraction

ProteinChip™
Surface Enhanced
Laser Desorption/
Ionization:
Time of Flight
Mass Spectrometry
SELDI-TOF-MS

CaP
Normal

Intensity

Molecular Weight

Center for Biomedical Proteomics
The George Washington University Medical School
Biomek® 2000 Laboratory Automation Workstation

Process 48 chips/384 samples per day
ProteinChip® Arrays: A Variety of Surfaces

Chemical Surfaces – Protein Expression Profiling:
- Hydrophobic
- Anionic
- Cationic
- Metal Chelate
- Normal Phase

Biological Surfaces – Protein Interaction Assays:
- PS-10 or PS-20
- Antibody - Antigen
- Receptor - Ligand
- DNA - Protein
ProteinChip® Array Preparation

**Step 1:** Complex protein sample is placed on a ProteinChip Array
- Proteins bind to chemical or biological sites on the ProteinChip surface

**Step 2:** Remove unbound proteins
- Wash the ProteinChip with appropriate stringency buffer
- Bound proteins are retained

**Step 3:** Add Energy Absorbing Molecules or “Matrix”
- EAM is applied to each spot to facilitate desorption and ionization in the TOF-MS Chip Reader
SELDI BioChip Arrays

Addressable locations
100 µm x 100 µm units
(10,000 per cm²)

Focused laser beam

Sample spot
Laser queries multiple Positions on sample spot

20
50
80
Laser Desorption/Ionization

Time-Of-Flight (TOF) Mass Spectrometry

Target: Matrix-embedded proteins on Chip

Laser

Detector

Flight Tube

Relative Intensity

analyte ions separated according to their mass/charge ratio

m/z
Data presentation options
An example of SELDI output

200 peaks for serum (upto 200K Da)

low mw

high mw
Pieces of the Puzzle

- Data acquisition, pre-processing, processing, signal versus noise
- Data interpretation phase 1, which signals are potential biomarkers
- Determining sources of variability, response to bias, analytical reproducibility
- Data interpretation phase 2, biomarker maturation
Flow diagram of spectrum analysis

1. SELDI-TOF MS
2. Baseline subtraction
   Normalization
3. Peak detection
4. Peak alignment and clustering
   Duplicates averaged
   80 peaks
5. Classification (Decision Tree)
SELDI Profiles of Normal vs. Cancer Serum Proteins
<table>
<thead>
<tr>
<th>Step 1: Discovery</th>
<th>Step 2: Evaluation</th>
<th>Step 3: Class prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training data set</strong></td>
<td><strong>Test data set</strong></td>
<td><strong>Unknown data set</strong></td>
</tr>
<tr>
<td>Disease</td>
<td>Normal</td>
<td>Disease</td>
</tr>
<tr>
<td>Profile 1</td>
<td>Profile 2</td>
<td>Profile 1</td>
</tr>
</tbody>
</table>

**Pattern discovery**

**Cluster analysis**

Use biomarker pattern for step 2.

**Determination of:**
- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value

---

*The George L. Wright, Jr.
Center for Biomedical Proteomics
Eastern Virginia Medical School*
Pieces of the Puzzle

- Data acquisition, pre-processing, processing, signal versus noise
- Data interpretation phase 1, which signals are potential biomarkers
- Determining sources of variability, response to bias, analytical reproducibility
- Data interpretation phase 2, biomarker maturation
Validation Team

- UTHSCSA
- EVMS
- UAB
- DMCC
- USUHS CPDR
- JHMI
- NCI FCRC
- UPCI
Table 2b  Inter-Lab variability

<table>
<thead>
<tr>
<th>Peak</th>
<th>Mass</th>
<th>Intensity</th>
<th>S/N</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak 1</td>
<td>average</td>
<td>5906.47</td>
<td>26.57</td>
<td>163.06</td>
</tr>
<tr>
<td></td>
<td>stdev</td>
<td>6.70</td>
<td>9.67</td>
<td>107.72</td>
</tr>
<tr>
<td></td>
<td>CV</td>
<td>0.0011</td>
<td>0.36</td>
<td>0.23</td>
</tr>
<tr>
<td>Peak 2</td>
<td>average</td>
<td>7768.61</td>
<td>35.94</td>
<td>242.75</td>
</tr>
<tr>
<td></td>
<td>stdev</td>
<td>8.41</td>
<td>6.25</td>
<td>82.77</td>
</tr>
<tr>
<td></td>
<td>CV</td>
<td>0.0010</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td>Peak 3</td>
<td>average</td>
<td>9289.18</td>
<td>30.96</td>
<td>244.03</td>
</tr>
<tr>
<td></td>
<td>stdev</td>
<td>9.89</td>
<td>4.70</td>
<td>77.35</td>
</tr>
<tr>
<td></td>
<td>CV</td>
<td>0.0011</td>
<td>0.15</td>
<td>0.18</td>
</tr>
</tbody>
</table>

SELDI-Prostate Cancer Validation

• **Phase I:** Synchronization of SELDI Instruments and Validation of Robotic Sample Processing: Can each site synchronize their instruments and procedures to match other sites?  
  *(STATUS: Successful Completion last year)*

• **Phase II:** Population Validation: Can multiple sites get the same result in a geographically diverse cross-section study?  
  *(STATUS: Initiating this year)*

• **Phase III:** Clinical Validation: Can multiple sites get the same result in a True Early Detection Analysis?
Pieces of the Puzzle

• Data acquisition, pre-processing, processing, signal versus noise
• Data interpretation phase 1, which signals are potential biomarkers
• Determining sources of variability, response to bias, analytical reproducibility
• Data interpretation phase 2, biomarker maturation
Summary of Biomarker Discovery and Identification

SELDI-TOF

Classification and Regression Tree Analysis

Validation

Identification

Purification

Western SDS-PAGE


The George L. Wright, Jr.
Center for Biomedical Proteomics
Eastern Virginia Medical School
Sample Preparation

1-Dimensional PAGE

- Remove Stain
- Dehydrate with 100% Acetonitrile
- Dry in Speed Vac
- Re-hydrate slice in 12.5 ng/µL Trypsin Solution
- Incubate 45 minutes on Ice
- Add Ammonium Bicarbonate to Keep Moist
- Incubate Overnight
- Extract with Formic Acid/Acetonitrile - Dry in Speed Vac
- Resuspend in Buffer “A”

2-Dimensional PAGE
For each scan in the Chromatogram, there are 3 “parent” ion masses each with an associated list of fragment ions:

<table>
<thead>
<tr>
<th>Peptide List (1046.7):</th>
<th>Parent Ion</th>
<th>Fragment Ions</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1029.4</td>
<td></td>
<td>16410.2</td>
<td></td>
</tr>
<tr>
<td>918.4</td>
<td></td>
<td>5211.5</td>
<td></td>
</tr>
<tr>
<td>836.6</td>
<td></td>
<td>6187.2</td>
<td></td>
</tr>
<tr>
<td>746.5</td>
<td></td>
<td>817.3</td>
<td></td>
</tr>
<tr>
<td>707.7</td>
<td></td>
<td>1181.3</td>
<td></td>
</tr>
<tr>
<td>673.7</td>
<td></td>
<td>10527.8</td>
<td></td>
</tr>
</tbody>
</table>

There are ~ 2500 scans/sample with a 90 min gradient
SELDI ProteinChip® Immunoassay

1. Capture/retention

+ sample

wash

Control Ab

Bait Ab

2. Detection

+ EAM

3. Data analysis

ApoA II Titration Curve

\[ y = 1.3198x \]

\[ R^2 = 0.985 \]

Mass/charge

Normalized Intensity

0 µg/ml

10 µg/ml

15 µg/ml

25 µg/ml

50 µg/ml

100 µg/ml

200 µg/ml

400 µg/ml

600 µg/ml

800 µg/ml

900 µg/ml

1000 µg/ml

0

100

200

300

400

500

600

700

800

ApoA II Conc. (ug/ml)
**SELDI Immunoassay.** A group of 50 NO (healthy), 51 BPH (benign) and 48 PCa (prostate cancer) samples were screened for ApoA-II levels using the SELDI-based immunoassay on mouse mAb Anti-ApoA-II coated PS20 ProteinChips®. The 8.9K m/z area observed in the serum samples was used to calculate the serum levels of ApoA-II based on the titration curve. About 4 fold increase in the protein amount in PCa samples is observed as compared to NO. (D) The relative normalized intensity observed in the SELDI-TOF-MS 8.9K m/z trace on IMAC-Cu2+ ProteinChips®.
ApoA II Immunodepletion

A

B

C

Average Intensity

<table>
<thead>
<tr>
<th></th>
<th>NO-Load</th>
<th>NO-IP</th>
<th>PCa-Load</th>
<th>PCa-IP</th>
<th>BPH-Load</th>
<th>BPH-IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg Intensity</td>
<td>10.887</td>
<td>1.588</td>
<td>60.406</td>
<td>11.921</td>
<td>74.164</td>
<td>8.621</td>
</tr>
</tbody>
</table>

The George L. Wright, Jr.
Center for Biomedical Proteomics
Eastern Virginia Medical School
**ApoA-II Immunohistochemistry.** Panel A: Prostate cancer infiltrating under uninvolved prostate glands (thick, double-headed gray arrow) which shows minimal staining even in basal cells. The pattern of staining, which is cytoplasmic, membranous and nuclear in cancer, is accentuated slightly on the advancing edge and is variable (weak to strong). Inset shows prostate cancer demonstrating scattered nuclear (thin, single-headed arrows) staining. **Panel B:** Prostate cancer infiltrating above uninvolved prostate glands (thick, double-headed gray arrow) which shows minimal focal staining. The pattern of staining in the prostate cancer is variable. **Panel C:** Weak to strong staining demonstrated in the luminal cells of PIN and prostate cancer (thin, single-headed arrows) present above uninvolved prostate glands with little staining (thick, double-headed gray arrow). Original magnification X400.
Ability of ApoA-II to discriminate controls from cases with-
Marginal Clinical Symptoms (DRE+/PSA- or DRE-/PSA+)
PSA<4.0, Biopsy Performed
• 46 Controls in which the biopsy was negative
• 46 Cases in which the biopsy was positive

Each sample was randomized in duplicate and analyzed
on antibody coated PS20 chips
In this range the ApoA-II may prove a useful biomarker for contributing to the PSA test by complementing this marker in the range where PSA fails to detect cancer.
WHAT NEXT?

• Facilitation of protein ID
  – LC-MS/MS
  – LIFT-MALDI TOF/TOF

• Fractionation, depletion and concentration steps
  – Higher resolution (MALDI TOF/TOF)
  – Upfront Fractionation
  – Depletion of abundant protein (IgY Fractionation)

• Sensitivity vs. resolution
  – Concentration of depleted serum
  – Improved instrument sensitivity

• Mass spec-assisted Immunoassay development
Bruker ClinProt Technology Process

1. Bind Protein Mixture to Magnetic Beads
2. Wash
3. Unbound
4. Bound
5. Elute Bound Protein

Intensities

m/z

Read in MALDI TOF/TOF

Spot eluted Proteins on AnchorPlate (396 spots)

MALDI Plate

The George L. Wright, Jr.
Center for Biomedical Proteomics
Eastern Virginia Medical School
203 total unique peaks mass range 1000-10000

WCX=84 peaks

IMAC=85 peaks

C18=62 peaks

WAX=80 peaks
MALDI profiling of serum IgY depletion

IgY unbound serum
1~10KDa 166 peaks

No IgY
1~10KDa 36 peaks
WHAT NEXT?

- Facilitation of protein ID
- Automated fractionation, depletion and concentration steps
- Sensitivity vs resolution
- Mass spec-assisted Immunoassay development
Development of these bioamarkers may provide important contributions to multiplexed immuno-assays or antibody arrays. The observed overexpression of these markers, like ApoA-II, in PCa patients with PSA < 4.0 ng, suggests that analysis of such biomarkers in serum may extend the utility of current blood testing for PCa.
Protein Expression Profiling with MS is reproducible, semi-quantitative, and can be validated.

MS-Profiling is an effective approach to biomarker discovery and has the potential to become a powerful analytical (diagnostic) tool.