Plasma Proteomics: Lessons in Biomarkers and Diagnostics

Leigh Anderson, PhD
leighanderson@plasmaproteome.org
leighanderson@SISCAPA.com

Exposome Workshop
Washington 8 Dec 2011
QUESTIONS AND LESSONS:

CLINICAL DIAGNOSTICS AS A MODEL FOR EXPOSOME INDICATORS

TECHNOLOGY OPTIONS FOR MEASURING PROTEIN RESPONSES TO EXPOSURES

SCALE OF THE PROBLEM: EXPOSURE SIGNALS VS POPULATION NOISE
The Clinical Plasma Proteome

- Plasma and serum are the dominant non-invasive clinical sample types
  - standard materials for *in vitro* diagnostics (IVD)

- Proteins measured in clinically-available tests in the US
  - 109 proteins via FDA-cleared or approved tests
    - Clinical test costs range from $9 (albumin) to $122 (Her2)
    - 90% of those ever approved are still in use
  - 96 additional proteins via laboratory-developed tests (not FDA cleared or approved)
  - Total 205 proteins (≈ products of 211 genes, excluding Ig’s)

- Clinically applied proteins thus account for
  - About 1% of the baseline human proteome (1 gene : 1 protein)
  - About 10% of the 2,000+ proteins observed in deep discovery plasma proteome datasets
Research Investments Have Generated Significant Scientific Output

Ptolemy and Rifai, Scandinavian Journal of Clinical & Laboratory Investigation, 2010; 70(Suppl 242): 6–14
Candidate Cancer Biomarkers Include ~25% of Human Proteins

Candidate List of your Biomarker (CLUB): A Web-based Platform to Aid Cancer Biomarker Research

<table>
<thead>
<tr>
<th>Candidate List of your Biomarker (CLUB): A Web-based Platform to Aid Cancer Biomarker Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernett Pardha N. Leigh</td>
</tr>
<tr>
<td>1Bioinformatics Institute, Florida</td>
</tr>
</tbody>
</table>

Abstract: The CLUB project is designed to support various cancers that have been rigorously studied. The different cancers examined have been identified in cancer biomarker literature. The CLUB project has used sets of data from various cancers and has combined them into a single database. The database contains the functions and other clinically relevant information.
“New” Protein Diagnostics Are FDA-Cleared at a Rate of ~1.5/yr: Insufficient to Meet Dx or Rx Development Needs

IT MAY BE AS DIFFICULT TO FIND A STRONG BIOMARKER AS IT IS TO FIND AN EFFECTIVE DRUG
A Major Technology Gulf Exists Between Discovery Proteomics and Routine Diagnostic Platforms: Virtuoso vs Pushbutton Analysis

<table>
<thead>
<tr>
<th>Discovery Proteomics: LC-MS/MS</th>
<th>Routine Clinical Analyzer: Immunoassay</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-700</td>
<td>1-20</td>
</tr>
<tr>
<td>$1,000-$10M</td>
<td>$2-100</td>
</tr>
<tr>
<td>25-50%</td>
<td>3-5%</td>
</tr>
<tr>
<td>4-52 wks</td>
<td>~15 min</td>
</tr>
<tr>
<td>2-50</td>
<td>100-1,000,000</td>
</tr>
<tr>
<td>Mix of auto &amp; manual steps</td>
<td>All inside 1 box</td>
</tr>
</tbody>
</table>
## Two Streams of Proteomics

<table>
<thead>
<tr>
<th>Problem</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic biology:</strong> maximum proteome coverage (including PTM’s, splices) to provide unbiased discovery of mechanistic information</td>
<td></td>
</tr>
<tr>
<td>- Critical: Depth and breadth</td>
<td><strong>Discovery proteomics</strong></td>
</tr>
<tr>
<td>- Not critical: Cost, throughput, quant precision</td>
<td>Specialized proteomics field, large groups, complex workflows and informatics</td>
</tr>
<tr>
<td><strong>Population biology:</strong> large sample numbers to detect statistical invariants, including biomarkers and clinical diagnostics</td>
<td></td>
</tr>
<tr>
<td>- Critical: Throughput, quant precision, depth and cost</td>
<td><strong>Directed assays</strong></td>
</tr>
<tr>
<td>- Not critical: Breadth</td>
<td>Primarily medical research, widely dispersed, needs simple workflows and data analysis</td>
</tr>
</tbody>
</table>
Protein Changes Identify and Distinguish Biological Mechanisms

The effects of peroxisome proliferators on protein abundances in mouse liver.

Anderson, N.L., Esquer-Blasco, R., Richardson, F., Foxworthy, P. and Eacho, P.
Toxicology and Applied Pharmacology, 137, 75-89, 1996.
Multivariate Protein Markers Resolve Drug Mechanisms
Data from Quantitative 2-D Gel Studies in Mouse Liver
With Test Panels Involving > 100 Proteins

Peroxisome Proliferator Treatments

A = Control
B = LY163,443
C = LY171,883
D = DEHP
E = Clofibric acid
F = WY14,643
G = Nafenopin

Each symbol represents the liver protein pattern for 100+ proteins in the liver of an individual mouse.

Peroxisome Proliferators: 6 Compounds Compared Over 107 Selected Protein Spots
The effects of peroxisome proliferators on protein abundances in mouse liver.
Anderson, N.L., Esquer-Blasco, R., Richardson, F., Foxworthy, P. and Eacho, P.
Toxicology and Applied Pharmacology, 137, 75-89, 1996.
Immunoassays Predominate for Clinical Protein Measurement

- Most proteins (80+%) are measured by sandwich immunoassay
  - ~$3B clinical test market
- Enzyme assays (17%) are legacy tests, of limited specificity but very low cost
- Functional assays primarily for coagulation (few immunoassays)

Technology Alternatives for Specific Assays for Protein Biomarker Studies

• Immunoassays (typical clinical test implementation)
  – Very sensitive
  – Expensive: IVD-quality assays cost $2-5 million
  – Specificity issues with less well-developed assays
  – Multiplexing limits in a single assay volume

• Hybrid MS-based assays
  – Peptide MS for quantitation and identification
  – Insensitive to folding, complexation, etc.
  – Absolute analyte specificity
  – Multiplex 25-200 assays/analysis
“Inside every bad protein is at least one good peptide”

- Tryptic peptides are effective analytical surrogates for proteins containing them
- A proteotypic peptide has a sequence that
  - Is unique in human proteome, as deduced from human genome sequence
  - May or may not be shared with other species (this can be checked for sequenced genomes and known recombinant proteins)
- A useful proteotypic peptide is characterized by
  - High yield upon tryptic digestion of sample
  - High signal strength in mass spectrometry by some ionization method (e.g., electrospray or MALDI)
  - No common variants or PTM’s, reasonable hydrophobicity & composition
- Tryptic digestion removes problems of IA
  - Eliminates post-acquisition enzymatic changes
  - Eliminates epitope-masking interferences
  - Permits facile creation of stable isotope labeled internal standards
Peptide-Level MS Provides High Structural Specificity
Multiple Reaction Monitoring (MRM) Quantitation

Q1 specific ion passed thru all others pulled away

Q0 All particles

Q2 specific ion fractured

Q3 specific mass sub particle allowed to pass

MRM

EQLGEFYEALDCLR^{++}
Mass = (1742.8+H^+)/2
= 871.9

FYEALDCLR^{++}
Mass = 1186.6

EQLGE

FYEALDCLR^{++}

FYEALDCLR^{+}

Mass = 1186.6
**MRM of Proteotypic Tryptic Peptides Provides Highly Specific Assays for Proteins > 1ug/ml in Plasma**

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Peptide Sequence</th>
<th>Protein Name</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afamin</td>
<td>DADPDUFFAK</td>
<td>Complement factor B</td>
<td>EELLPAQDIK</td>
</tr>
<tr>
<td>Albumin</td>
<td>LVNEVTEFAK</td>
<td>Complement factor H</td>
<td>SPDVINGSPISQK</td>
</tr>
<tr>
<td>α-1-acid glycoprotein</td>
<td>NWGSLSVYADKPETTK</td>
<td>Fibrinogen α</td>
<td>GSESGIFTNTK</td>
</tr>
<tr>
<td>α-1-antichymotrypsin</td>
<td>EIGELYLPK</td>
<td>Fibrinogen β</td>
<td>QGFGNVATNTDGK</td>
</tr>
<tr>
<td>α-1B-glycoprotein</td>
<td>LETPDFQLFK</td>
<td>Fibrinogen γ</td>
<td>DTVQIHDITGK</td>
</tr>
<tr>
<td>α-2-macroglobulin</td>
<td>LIGNQEPGGQTKALK</td>
<td>Fibronecin</td>
<td>DLQFVEVTVDVK</td>
</tr>
<tr>
<td>Angiotensinogen</td>
<td>PKDPTFIPAPIQAK</td>
<td>Gelsolin, isoform 1</td>
<td>TGAQELLRL</td>
</tr>
<tr>
<td>Antithrombin-III</td>
<td>DDLYVSDAFHK</td>
<td>Haptoglobin β</td>
<td>VGYVSGWGR</td>
</tr>
<tr>
<td>Apolipoprotein A-I</td>
<td>ATEHLSTLSEK</td>
<td>Hemopexin</td>
<td>NFPSPVDAAFR</td>
</tr>
<tr>
<td>Apolipoprotein A-II</td>
<td>SPELQAEAK</td>
<td>Heparin cofactor II</td>
<td>TLEAQLTPR</td>
</tr>
<tr>
<td>Apolipoprotein A-IV</td>
<td>SLAPYQADTQEK</td>
<td>Histidine-rich glycoprotein</td>
<td>DSPVLIDFFEDTER</td>
</tr>
<tr>
<td>Apolipoprotein B-100</td>
<td>FPEVDVLTK</td>
<td>Inter-α-trypsin inhibitor HC</td>
<td>AAISGENAGLVR</td>
</tr>
<tr>
<td>Apolipoprotein C-I</td>
<td>TPDVSSALDK</td>
<td>Kininogen</td>
<td>TVGSSTFYSFK</td>
</tr>
<tr>
<td>Apolipoprotein C-III</td>
<td>DALSSQESQVAQQAR</td>
<td>L-selectin</td>
<td>AEIEYLEK</td>
</tr>
<tr>
<td>Apolipoprotein E</td>
<td>LGPLVEQGR</td>
<td>Plasma retinol-binding protein</td>
<td>YWGVASFLQK</td>
</tr>
<tr>
<td>β-2-glycoprotein I</td>
<td>ATVVYQGER</td>
<td>Plasminogen</td>
<td>LFLEPTR</td>
</tr>
<tr>
<td>Ceruloplasmin</td>
<td>EYTDASFNTNR</td>
<td>Prothrombin</td>
<td>ETAASLLQAGYK</td>
</tr>
<tr>
<td>Clusterin</td>
<td>LFDSDPITVTVPVEVS</td>
<td>Serum amyloid P</td>
<td>VGEYSLYGR</td>
</tr>
<tr>
<td>Coagulation factor XIIa HC</td>
<td>VVGGGLVALR</td>
<td>Transferrin</td>
<td>EDPQTFYYYAVVK</td>
</tr>
<tr>
<td>Complement C3</td>
<td>TQLQEVEVK</td>
<td>Transthyretin</td>
<td>AADTWEFPASGK</td>
</tr>
<tr>
<td>Complement C4 β</td>
<td>ITQVLHFTK</td>
<td>Vitamin D-binding protein</td>
<td>THLPEVFLSK</td>
</tr>
<tr>
<td>Complement C4 γ</td>
<td>VGDTLNRLR</td>
<td>Vitronectin</td>
<td>FEDGVLDPDPYPR</td>
</tr>
<tr>
<td>Complement C9</td>
<td>AEEDYINEFSVR</td>
<td>Zinc-α-2-glycoprotein</td>
<td>EIPAWVPFDPAQITK</td>
</tr>
</tbody>
</table>

ADDRESSING MRM LIMITATIONS:
SENSITIVITY
THROUGHPUT (LC-MS/MS CYCLE TIME)
SISCAPA: Enrich Target Peptides and Decrease Sample Complexity

Direct injection of unfractionated plasma digest

10 nl plasma →

SISCAPA enrichment of targeted peptides

Ab to target peptide → wash →

10 ul plasma →

1000x larger digest input volume + Reduced ion suppression + Same LC-MS/MS = >1,000-fold increased MRM assay sensitivity
10-Plex SISCAPA Capture
(10 x 1 µg pAb + 10 µl plasma equiv digest)

Target peptides are enriched (from near baseline) and other plasma peptides (Alb, Hp) are depleted
Standard Curves Characterizing SISCAPA Assay for Soluble Mesothelin Peptide LLGPHVEGLK in Plasma

- Standard addition curve (**Fwd**) indicates endogenous level of 5fmol/10ul plasma
  - 18 ng/ml protein
  - 2% average within-run CV
- IS dilution curve (**Rev**) indicates LOQ of ~50amol/10µl plasma
  - 200 pg/ml protein
  - 8% average within-run CV
The Zolg Number (Z*)

- Clinically believable biomarkers can emerge when it is practical to:
  - produce good measurements of candidates (CV<15%)
  - in Z samples, where Z is a disease specific number, typically \( \approx 1,500 \), comprising multiple disease groups and controls
- Of the 20,075 publications on “protein biomarkers”, I know only two that used >1,000 samples
- With sample numbers <Z, substantial doubt remains as to validity in human populations
- If 1,000+ samples can be run in a week, biomarker studies can become realistic

* Attributed to Werner Zolg, formerly of Roche Diagnostics
TRANSLATING BIOMARKER ASSAYS TO THE CLINIC: STEPS TOWARDS REGULATED USE OF HIGH-THROUGHPUT PROTEIN ASSAYS
Thyroglobulin is used as a marker for recurrence of thyroid cancer.

Known interferences prevent use of existing clinical immunoassays in 20-40% of patients.

SISCAPA assay eliminates these interferences.

In collaboration with Dr. Hoofnagle SAT is providing improved SISCAPA antibodies that bring test sensitivity to clinically required level.
Protein-Based Multiplex Assays: Mock Presubmissions to the US Food and Drug Administration

Fred E. Regnier,1 Steven J. Skates,2 Mehdi Mesri,3 Henry Rodriguez,3 Živana Težak,4 Marina V. Kondratovich,4 Michail A. Alterman,5 Joshua D. Levin,4 Donna Roscoe,4 Eugene Reilly,4 James Callaghan,4 Kellie Kelm,4 David Brown,6 Reena Philip,4 Steven A. Carr,7 Daniel C. Liebler,8 Susan J. Fisher,9 Paul Tempst,10 Tara Hiltke,3 Larry G. Kessler,11 Christopher R. Kinsinger,3 David F. Ransohoff,12 Elizabeth Mansfield,4 and N. Leigh Anderson13*
A Human Proteome Detection and Quantitation Project*


The lack of sensitive, specific, multiplexable assays for most human proteins is the major barrier impeding development of candidate biomarkers into clinically useful tests. Recent progress in mass spectrometry-based assays for proteotypic peptides, particularly those with specific affinity peptide enrichment, offers a systematic and economical path to comprehensive quantitative coverage of the human proteome. A compete suite of assays, e.g. two peptides from the protein product of each of the ~20,500 human genes (here termed the human Proteome Detection and Quantitation project), would enable rapid and systematic verification of candidate biomarkers and lay a quantitative foundation for subsequent efforts to define the larger universe of splice variants, post-translational modifications, protein-protein interactions, and tissue localization. *Molecular & Cellular Proteomics 8: xxx-xxx, 2009.

benefits of taking time to involve the entire biological research community and especially the medical research segment, in these discussions are substantial.

Progress in systematically measuring proteins, however, need not wait for the conclusion of such discussions. We propose a near-term tactical approach, called the human Proteome Detection and Quantitation (hPDQ) project that will enable measurement of the human proteome in a way that would yield immediately useful results whereas the strategy for a comprehensive Human Proteome Project is worked out. The hPDQ project is aimed at overcoming present difficulties in answering basic biological questions about the relationship between protein abundance (or concentration) and gene expression, phenotype, disease, and treatment response; i.e. the growing field of protein biomarkers. It is thus focused on the study of biological variation affecting protein expression.
Acknowledgments

Academic Partners and Funding

• University of Victoria, BC
  – Terry Pearson, Matt Pope, Angela Jackson, Morteza Razavi, Department of Biochemistry and Microbiology, University of Victoria, B.C, Canada
  – Christoph Borchers, Derrick Smith, Darryl Hardy, UVic-Genome B.C. Proteomics Centre

• Carr CPTAC Team
  – Steve Carr, Eric Kuhn, Terri Addona, Hazmik Keshishian, Sue Abbatiello, Broad Institute
  – Mandy Paulovich, Jeff Whitaker, Lei Zhao, Fred Hutchinson Cancer Research Center

• Grant Funding
  – National Cancer Institute
    • Biomarker Discovery Initiative (contract # 23XS144A)
    • Clinical Proteomic Technology Assessment for Cancer (grant U24-CA126476-01)
  – Canary Fund Seed Grant

Corporate Partners

• Epitomics: RabMAb Antibody Development
  – Guoliang Yu, Xiuwen Liu, Taiying Chen

• Agilent: SISCAPA Platform Optimization
  – Chris Miller, Keith Waddell, Gus Salem, Agilent Mass Spectrometry
  – Marc Beban, Peter Werner, Nitin Sood, Agilent Automation Solutions

• ABSciex: Plasma MRM Assays
  – Christie Hunter, Tina Settineri, Applied Biosystems, Foster City

• ThermoFisher: Kingfisher

• Assay Collaborations
  – Pfizer
  – TGEN
  – biOasis
  – Mayo Clinic
  – Andy Hoofnagle, U of Washington

• Animation and Graphics
  – Beth Anderson, Arkitek Studios, Seattle

Commercialization of SISCAPA® technology is undertaken by SISCAPA Assay Technologies, Inc. www.SISCAPA.com
SISCAPA Publications


- **SISCAPA Peptide Enrichment on Magnetic Beads Using an Inline Beadtrap Device.** N. Leigh Anderson, Angela Jackson, Derek Smith, Darryl Hardie, Christoph Borchers, and Terry W. Pearson, Mol Cell Proteomics 8:995-1005 (2009)


- **MALDI Immunoscreening (MiSCREEN): A Method for Selection of Anti-peptide Monoclonal Antibodies For Use in Immunoproteomics.** Matthew E. Pope, Martin V. Soste, Brett A. Eyford, N. Leigh Anderson and Terry W. Pearson, in press