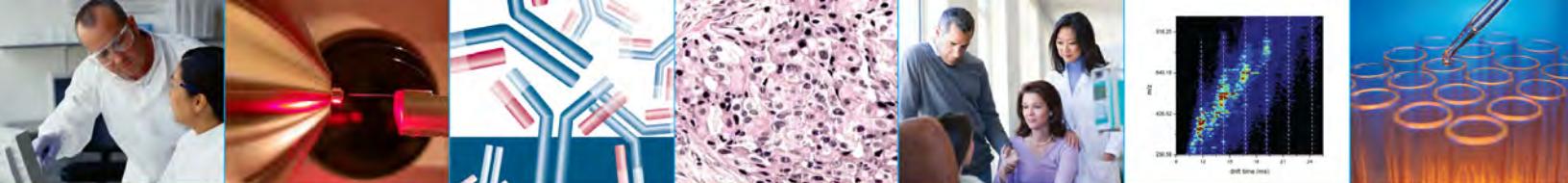


# Clinical Proteomic Technologies for Cancer

***Advancing Protein Science for Personalized  
Medicine***

U.S. Department of Health and Human Services  
National Institutes of Health  
National Cancer Institute



## EXECUTIVE SUMMARY

The addition of protein biomarker panels to the cancer diagnostic armamentarium is an area of considerable interest in medicine. The discovery that proteins and peptides are “leaked” by tumors into clinically accessible bodily fluids such as blood and urine has led to the possibility of diagnosing cancer at an early stage or monitoring response to treatment simply by collecting these fluids and testing for the presence of cancer-related biomarkers. Prostate-specific antigen (PSA) and cancer antigen 125 (CA-125) are examples of blood-borne cancer protein biomarkers that are currently being used in the clinic. However, the measurement of individual biomarkers has clinical limitations with respect to both sensitivity and specificity. For this reason, combinations of protein/peptide analytes are under intense investigation as biomarker panels can potentially bring greater sensitivity and specificity to cancer screening than any one analyte alone. If the potential of personalized medicine is to be realized, it must include this next generation of molecular diagnostics.

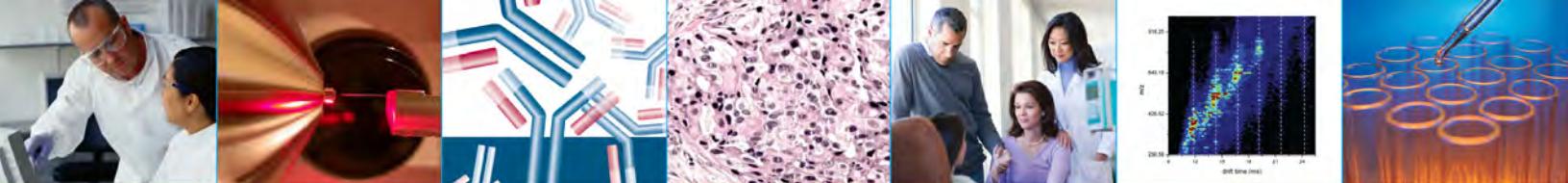
We do not suffer a lack of candidate protein biomarkers. As of 2006, there were 1,261 putative cancer protein/peptide biomarkers described in the scientific literature. The sobering reality, however, is that very few of these candidates have been validated, and even fewer have made it into a medical diagnostic product (Figure 1). This discrepancy indicates that the issue lies within the candidate biomarker pipeline.

**Figure 1. Reality Check**

Adapted from Ludwig and Weinstein (Nov. 2005). Biomarkers in Cancer Staging, Prognosis and Treatment Selection. *Nature Rev Cancer* 5, 845-856.



Proteomic technologies, which hold great promise for the discovery of novel cancer biomarkers, can help make sense of the complexities inherent in the proteome. In recent



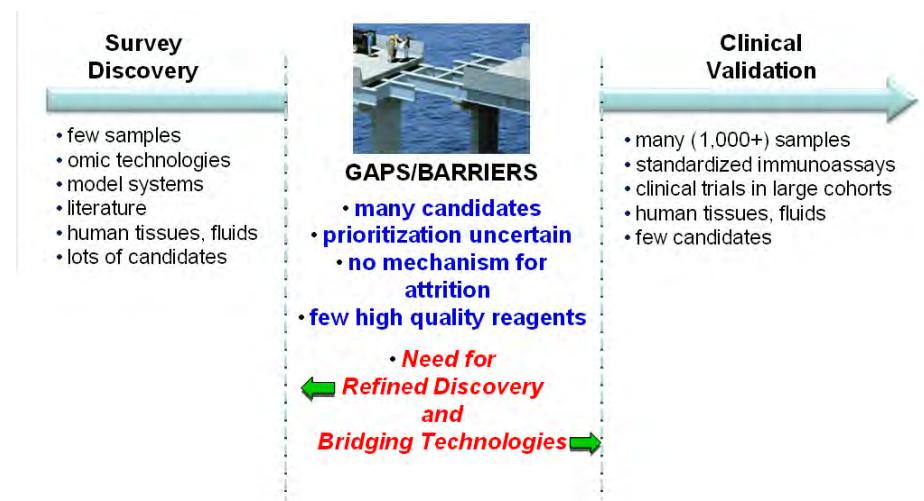
years, however, studies that have applied proteomic technologies to clinical applications—such as mass spectrometry and affinity-based detection methods—have met with some disappointment.

There are two major issues at hand:

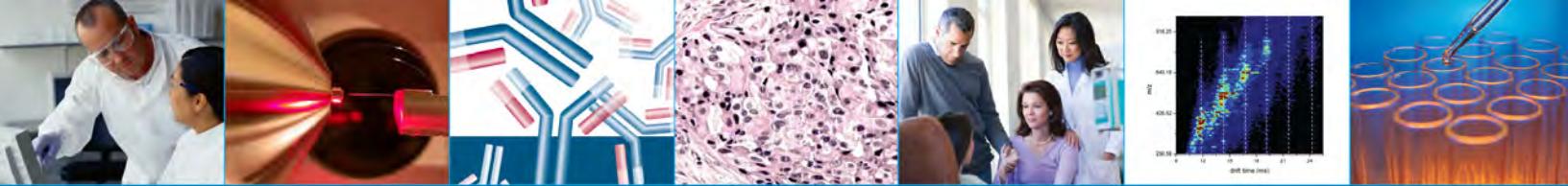
**1) Variability within biomarker discovery.** A paucity of standard reagents and methods for protein identification and measurement has led to pervasive problems with reproducibility and comparison of research results among laboratories, posing a significant challenge to the translation of discoveries to clinical applications.

**2) Biomarker candidates need to be pre-validated, or verified, prior to costly clinical validation studies.** For clinical validation of protein biomarkers, an enzyme-linked immunosorbent assay (ELISA) is developed for each antigen to test large cohorts in clinical trials. However, these tests are very expensive. Each ELISA takes up to one year and \$2 million to develop. A more efficient biomarker development pipeline will require coupling biomarker discovery in tissue and proximal fluid to verification in plasma *prior to* clinical validation (Figure 2). Verification is a rapid way to assess if a given candidate is detectable in blood and changes in a measurable way in relation to the presence or stage of disease. This bridging technology can rapidly triage a lengthy list of candidates prior to investing very large sums of money and time on the development of antibodies suitable for use in an ELISA.

**Figure 2. A Better Bridge is Needed between Biomarker Discovery and Clinical Validation**



Traditional approaches have contributed all they can; it is time for new technologies and approaches. To make clinical cancer proteomics a reality, the scientific community must



first invest in much needed technologies and infrastructure in order to build a better biomarker development pipeline. The goal of the Clinical Proteomic Technologies for Cancer (CPTC) initiative is to develop a more refined, efficient and reliable biomarker development pipeline. This pipeline is anticipated to produce better credentialed candidate leads, ultimately accelerating the translation of new cancer biomarkers into diagnostic tests.

Fixing this pipeline is too great an endeavor for a single investigator or institution. Accordingly, CPTC has brought together the best minds in proteomics to accomplish this goal. Together, the CPTC network is laying the foundation for clinical cancer proteomics by addressing each of the following barriers:

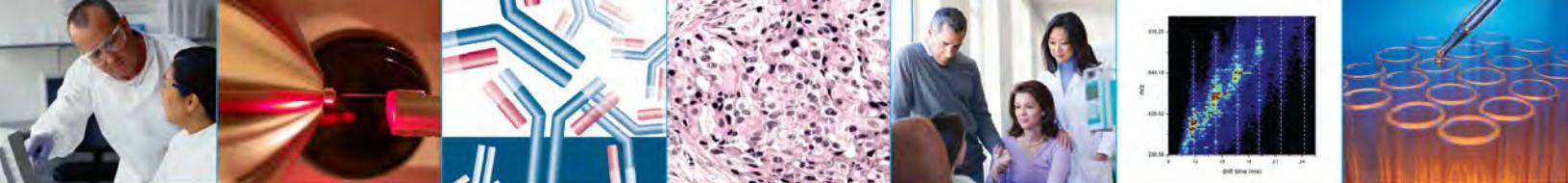
**Optimizing proteomic technologies and developing appropriate standards.** Current and emerging protein measurement technologies must be optimized and calibrated through the use of standard protocols and performance reagents to produce comparable results among laboratories.

**Standardizing procedures for collecting, processing, and storing biological samples used in proteomics research. The use of high-quality biospecimens is critically important for proteomic research because the output—the data—is only as good as the input.** The methods of biological sample preparation must be made more consistent to reduce variability in experimental results. Uniform sample quality, as well as access to large numbers of high-quality samples, will lead to more reliable results.

**Making high-quality reagents available and accessible.** These include capture reagents (e.g., antibodies) that can be used in protein arrays, as well as other techniques that are used to measure proteins. However, a key challenge for proteomic researchers is acquiring high-quality, well-characterized monoclonal antibodies. While numerous commercial reagent suppliers make antibodies available for research, they tend to be expensive and may or may not be extensively characterized.

**Developing technologies that can quantify proteins across a large dynamic range.** Enormous variation in protein concentrations and modifications are found within cells and body fluids, and the development of ultra-sensitive detection methods that have a large dynamic range has been a significant challenge. Capturing and identifying dilute proteins from a complex mixture has proven to be especially difficult. A biomarker present at 10 pg/mL in human plasma is roughly equivalent to trying to locate a softball somewhere between earth and the moon! In addition, many of the best techniques for finding low abundance proteins are not quantitative—that is, they can determine whether the protein is present or not but they cannot measure its concentration.

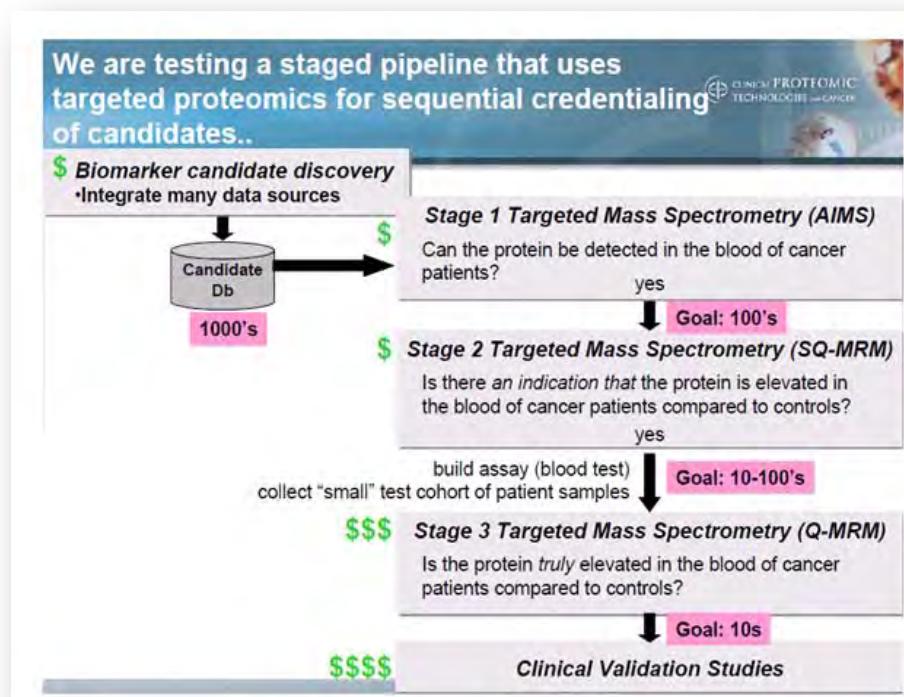
**Developing common bioinformatics resources with shared algorithms and standards for processing, analyzing, and storing proteomic data.** Proteomic informatics tools that permit data sharing and computation among laboratories are essential for rapid progress in the field.



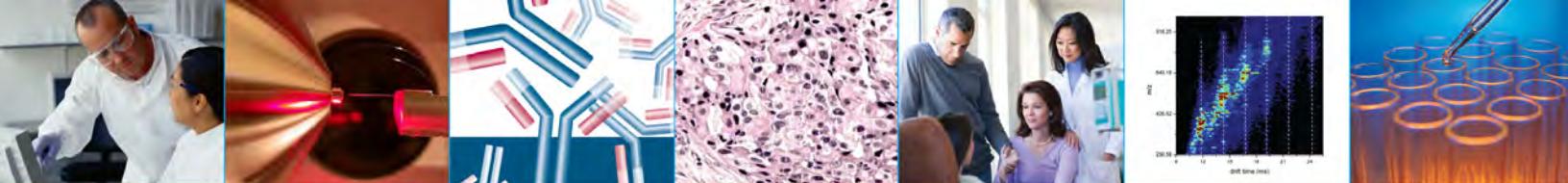
**Implementing a verification step in the protein biomarker pipeline.** Protein biomarker candidates should be “pre-qualified” before costly clinical validation studies are initiated. This important step can be accomplished by implementing a verification stage in the biomarker pipeline (Figure 3).

**Figure 3. Verification: Pre-qualifying biomarker candidates prior to costly clinical validation**

*For clinical validation of protein biomarkers, an ELISA is developed for each antigen in order to test large cohorts in clinical trials, which can take up to one year and millions of dollars to develop. A more efficient biomarker development pipeline will triage candidates before an ELISA is developed, providing greater confidence in candidates prior to costly clinical validation. Using targeted proteomics, CPTC is creating this bridge from discovery to validation, which will allow investigators to run up to 200 assays in a matter of months for a fraction of the cost of a single ELISA.*



**Adopting an interdisciplinary team approach to science.** No one laboratory working on its own could possibly examine all of the candidate biomarkers, develop all of the necessary technologies, or assemble all of the pieces of evidence required to



understand the molecular mechanisms of disease. It will require many laboratories working together to accomplish these goals.

In many ways, the challenges facing the clinical proteomic community are comparable to those that were faced by the genomics community prior to the Human Genome Project (HGP)—the current technology enables the sampling of only a small portion of the proteome and at different levels of quality. Visionaries brought the HGP to life, but it was improvements made in DNA sequencing technologies that made this endeavor possible. The technologies became high-throughput, reliable, and reproducible. Had the project moved forward without these much needed technological improvements, the genomics community would never have been able to live up to its promise and the opportunities afforded by the HGP would never have been realized. Clinical proteomics stands at a juncture: Do we improve the technologies and methodologies first—do we forge ahead or build a modern highway that can routinely deliver biomarkers to the cancer research community? CPTC is building the highway.

This document is a report on the first two years of progress in this five-year program. Some high level accomplishments include:

#### Biospecimens:

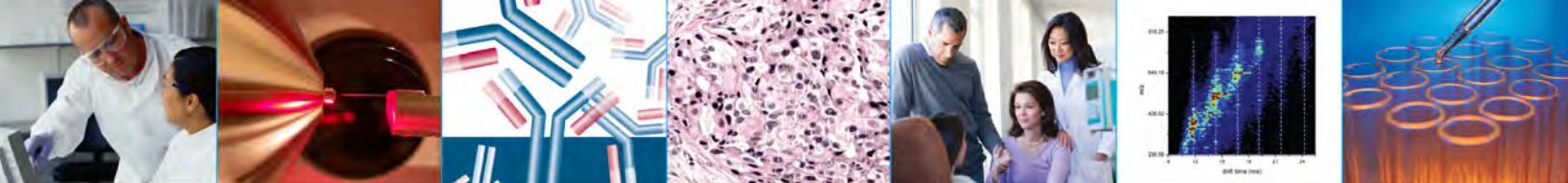
- A large collection of breast cancer cases (expected N=500) and control plasma samples (expected N=1500) have been accrued prior to diagnosis, to avoid bias of "baseline inequality."

#### Discovery Technologies for Proteomics:

- **Assessment of discovery proteomics technology platforms across laboratories.** The first large-scale evaluation of proteomics platforms for unbiased discovery of biomarker candidates has been completed. These studies defined reproducibility and sensitivity for detection of biomarker candidate proteins at defined levels in complex proteomes.
- **Development of standard proteomes for technology assessment.** A yeast reference proteome extract and defined protein standard spikes and mixtures, together with multilaboratory datasets, provide the community with resources for system calibration and standardization.
- **Development of performance metrics for discovery proteomics technology platforms.** A performance metrics “toolkit” to monitor performance of system components enables targeted optimization and troubleshooting of systems and long-term quality control.

#### Verification Technologies for Proteomics:

- **Process to identify candidate markers.** The first large-scale evaluation of targeted MS technology, Multiple Reaction Monitoring Mass Spectrometry (MRM-MS), has been completed for sorting through large lists of biomarker candidates to identify the most promising ones to advance to clinical validation.
- **Development of MRM-MS technology.** MRM-MS technology, because of its

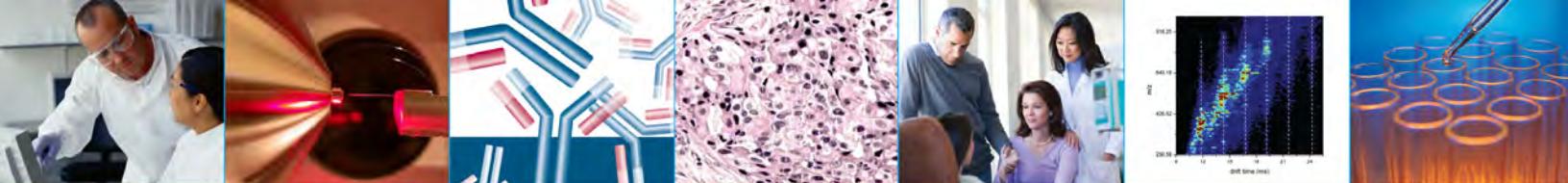


ability to measure low abundance proteins with a high degree of specificity, has the potential to serve as the critical filter to assess protein candidate performance in the absence of immunoassays.

- **Assessment of MRM-MS technology in complex biological mixtures.** MRM-MS technology has been tested in the context of unfractionated plasma, the most complex biological matrix of all. We have demonstrated that multiplexed, quantitative MS-based assays can be rapidly and robustly configured and deployed across laboratories for measurement of peptides in plasma, and that near-clinical assay performance with respect to reproducibility can be achieved.
- **Development of reagent libraries and standards as community resources.** The team has developed sets of reagents, methods and large, high quality datasets all of which will be available to the community. These materials will aid acceptance and adoption of MRM-MS technology by proteomics and clinical communities.
- **Dissemination of methods.** A training course is being developed to further promulgate the methods.

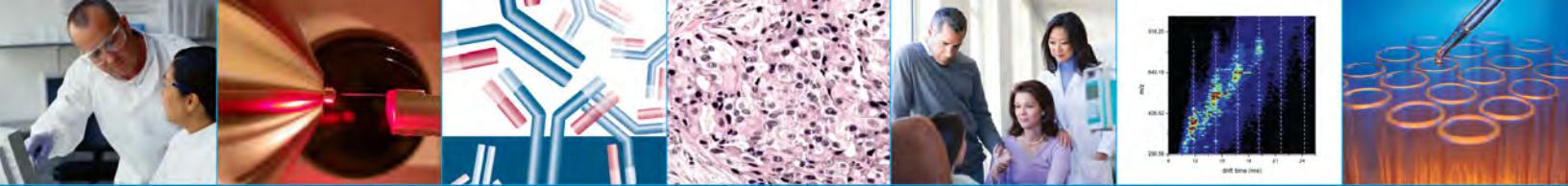
We are particularly gratified by the tremendous progress already made, which is a direct reflection of both the dedication to the highest quality and standards by all the CPTC members, and the deep commitment to open and collaborative science for the sake of the entire cancer proteomics community. Their work will have implications far beyond cancer proteomics, but their most lasting legacy will be the impact it will have on reducing the burden of suffering and death due to cancer. This is the ultimate reason we are all tirelessly working to ensure the success of this program.

CPTC Program Coordinating Committee

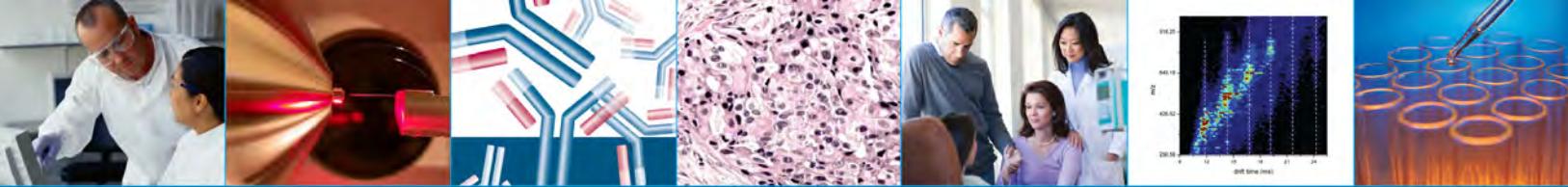


## CONTENTS

<b>Executive Summary</b>	2
<b>1. Program Operation</b>	<b>11</b>
1.1. Introduction	11
1.2. Program Infrastructure	12
1.2.1. Clinical Proteomic Technology Assessment Centers (CPTACs)	12
1.2.2. Advanced Platforms and Computational Sciences Awards	18
1.2.3. The Proteomic Reagents and Resources Core	20
1.3. Governance	21
1.3.1. Program Coordinating Committee	24
1.3.2. Communications Plan	24
1.4. Program Management	24
1.4.1. Scientific Working Groups	24
1.4.2. Site Visits	24
1.4.3. Informatics and caBIG Connectivity	25
1.4.4. Supporting Partnerships	26
1.4.5. Software Commercialization	26
1.4.6. Biospecimens and NCI Best Practices	28
1.4.7. Communication and Working with Outside Organizations	29
1.4.7.1. Communication Materials	29
1.4.7.2. Advocacy Outreach Activities	32
1.4.7.3. Annual Investigators' Meeting	38
1.4.7.4. NCI-NIST Next-Generation Methods of Peptide Identification Workshop	39
1.4.7.5. International Proteomics Data Release and Sharing Policy Summit (Amsterdam Principles)	40
1.4.7.6. Alignment with FDA Critical Path	41
1.4.7.7. NCI-FDA Interagency Oncology Task Force Molecular Diagnostics Workshop	41
1.4.7.8. National Heart, Lung and Blood Institute	43
1.4.7.9. American Association for Clinical Chemistry	43
1.4.7.10. Korean Functional Proteomics Center	44
1.4.7.11. Biotechnology and Pharmaceutical Industry	44
1.4.7.12. Institute of Medicine of the National Academies	45
1.4.7.13. American Association of Cancer Research (AACR)	46
1.4.7.14. Human Proteome Organization (HUPO)	46
1.5. Leveraged Funding	47
1.6. Public-Private Partnerships	50
1.6.1. Small Business Innovation Research	50
1.6.2. Innovative Molecular Analysis Technologies	51
1.6.3. University of Iowa Development Studies Hybridoma Bank	52



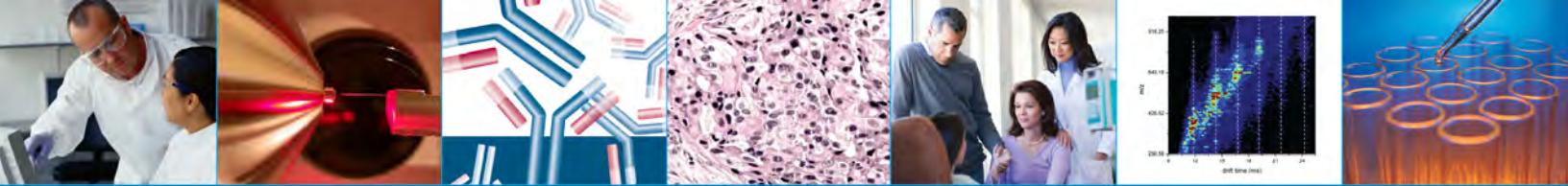
<b>2. Program Performance</b>	53
2.1. Introduction	53
2.2. Mass Spectrometry Platform Benchmarking “Round Robin Evaluations”	54
2.2.1. Multiple Reaction Monitoring Performance Assessment	54
2.2.2. A Reference Yeast Proteome for LC-MS/MS Performance	59
2.2.3. Performance Metrics for Evaluating LC-MS/MS Systems	61
2.2.4. Repeatability and Reproducibility in LC-MS/MS Analysis	62
2.3. Mass Spectrometry Platforms	64
2.3.1. Existing and Emerging Discovery Platforms	64
2.3.1.1. Shotgun Proteomics Technology Platforms for Unbiased Discovery	64
2.3.1.2. Assessment of Serum Peptide Profiling to Detect Cancer Specific Patterns	69
2.3.1.3. Targeted and Global Proteomic Strategies for Early Breast Cancer Detection	75
2.3.1.4. A Proteomics Platform for Quantitative, Ultra-High Throughput, and Ultra-Sensitive Biomarker Discovery	80
2.3.2. Targeted Platforms	83
2.3.2.1. Measuring Cancer Biomarker Candidates by Targeted MS and Antibody Enrichment	83
2.4. Mass Spectrometry Platforms for Post-Translational Modification Characterization	89
2.4.1. Proteomic Characterization of Alternate Splicing and cSNP Protein Isoforms	89
2.4.2. A New Platform to Screen Serum for Cancer Membrane Proteins	91
2.4.3. Quantitative Methods for Spectral and Image Data in Proteomics Research	93
2.4.4. A Proteomics Approach to Ubiquitination	95
2.4.5. A Cell-Free System for High-Yield Phosphoprotein Synthesis	97
2.4.6. CPTAC Post-Translational Modification Workflows	98
2.5. Immunoaffinity Platforms and Separation Technologies	99
2.5.1. Analytical Proteomics	99
2.5.2. Automated multiplexed immunoassays for rapid quantification of low abundance cancer-related proteins	106
2.5.3. Multiplex Mass Spectrometric Immunoassays	108
2.5.4. Immunoaffinity Capture Coupled with Ion Mobility	109
2.5.5. Highest sensitivity cancer marker array on Quadraspec’s Bio-CD platform	110
2.5.6. Automated multi-array platform for cancer biomarkers	111
2.5.7. Sensitive Protein Detection Combining Mass Spectrometry and PCR	112
2.5.8. High-throughput Selection of Aptamers Against Cancer Biomarkers	114
2.5.9. Proteomic Phosphopeptide Chip Technology for Protein Profiling	115
2.5.10. Aptamer-Based Proteomic Analysis for Cancer Signatures	118
2.5.11. Top-Down Mass Spectrometry of Salivary Fluids for Cancer Assessment	119
2.6. Reagents and Resources Development	121



2.6.1. CPTC Reagents and Resources Core	121
2.6.11 Reagents	122
2.6.1.2 Resources	130
2.6.2. Global Production of Disease Specific Monoclonal Antibodies	137
2.6.3. Developing Synthetic Antibodies for Array-based Cancer Detection	139
2.6.4. Yeast Single Chain Antibodies as Capture Reagents	140
2.6.5. Analysis and Statistical Validation of Proteomic Datasets	142
2.6.6. Standard Reference Materials	144
2.6.7. A Platform for Pattern-Based Proteomic Biomarker Discovery	146
2.6.8. PICquant-An Integrated Platform for Biomarker Discovery	148
2.6.9. New Proteomic Algorithms to Identify Mutant or Modified Proteins	150
2.6.10. Enhancement of MS Signal Processing Toward Improved Cancer Biomarker Discovery	152
2.6.11 Computational Tools for Cancer Proteomics	155

### **3. Training Opportunities in Emerging Technologies** 159

<b>4. Appendix</b>	<b>160</b>
4.1. Organizations Participating in the CPTC Initiative	160
4.2. Applied Proteomic Platforms and Computational Sciences Investigator Biographies	161
4.3. Responsibilities of CPTAC Members	167
4.4. Program Coordinating Committee Biographies	170
4.5. CPTAC Working Group Membership	177
4.6. Site Visit Guidelines and Checklist	180
4.7. eProtein Newsletters	
4.8. Annual Investigator's Meeting Agendas	
4.9. NCI-FDA Memorandum of Understanding	
4.10. NCI-FDA Interagency Task Force Molecular Diagnostics Workshop Agenda	
4.11. NHLBI Adoption of CPTAC Verification Technology	
4.12. Trade Media Article Reprints	
4.13. CPTC Publications	
4.14. CPTC Patent Disclosures/Applications Listing	
4.15. Upcoming and Past Speaking Engagements for CPTC Leadership	
4.16. CPTC Program Staff Biographies	



## 1. Program Operation

### 1.1 Introduction

#### Mission:

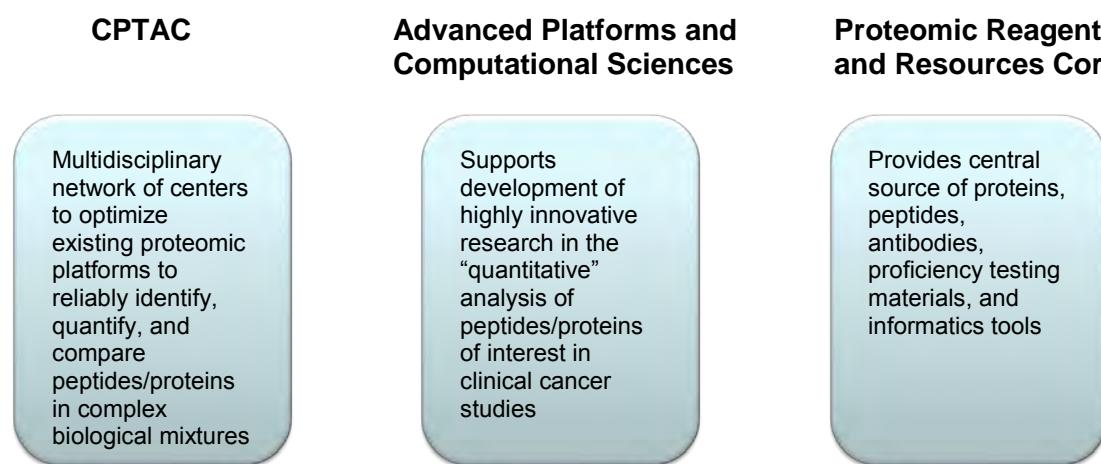
The NCI Clinical Proteomic Technologies for Cancer (CPTC) initiative seeks to foster the building of an integrated foundation of proteomic technologies, data, reagents and reference materials, and analysis systems to systematically advance the application of protein science to accelerate discovery and clinical research in cancer.

#### Goals:

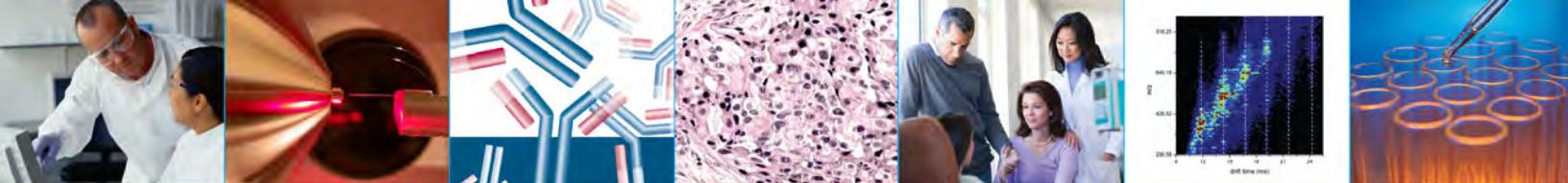
- Enhance technical abilities to identify and measure proteins accurately and reproducibly in biological systems.
- Advance proteomics as a reliable, quantitative field that can accelerate discovery and translational research.

The management of the CPTC initiative is an exceedingly complex endeavor. There are three integrated CPTC programs designed to overcome barriers to the application of proteomic technologies to clinical cancer research (Figure 4): the Clinical Proteomic Technology Assessment for Cancer (CPTAC) program, the Proteomic Reagents and Resources Core, and the Advanced Platforms and Computational Sciences program.

**Figure 4. CPTC Integrated Program Components**

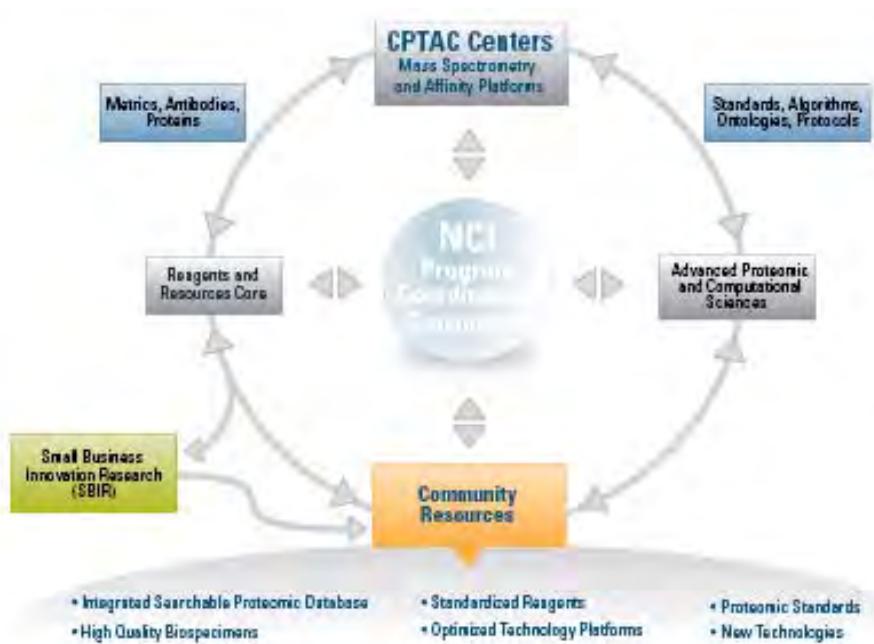


The CPTC infrastructure consists of scientists from nearly 50 federal, academic, and private sector organizations (Appendix 4.1). Researchers funded through the program represent seasoned and senior investigators who are leading large centers and multi-project efforts as well as junior faculty involved in the individual projects. At the heart of



the program are CPTAC sites (CPTAC Network) that have been awarded as cooperative agreements (U24s). The Program Office maintains close interaction with the investigators and closely monitors progress. Furthermore, the program staff promotes and monitors inter-network collaborations among different centers. The NCI program management occurs in conjunction with the operation of the Program Coordinating Committee (PCC), an oversight body that gathers individuals from the centers and also has representation from NCI (Figure 5).

**Figure 5. CPTC Program Management**

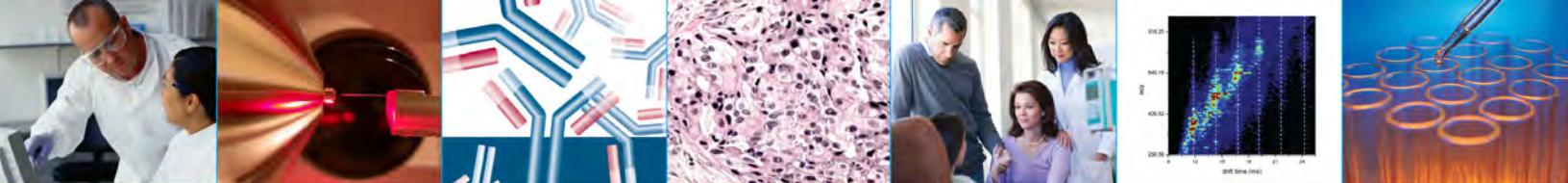


## 1.2 Program Infrastructure

### 1.2.1 Clinical Proteomic Technology Assessment Centers (CPTACs)

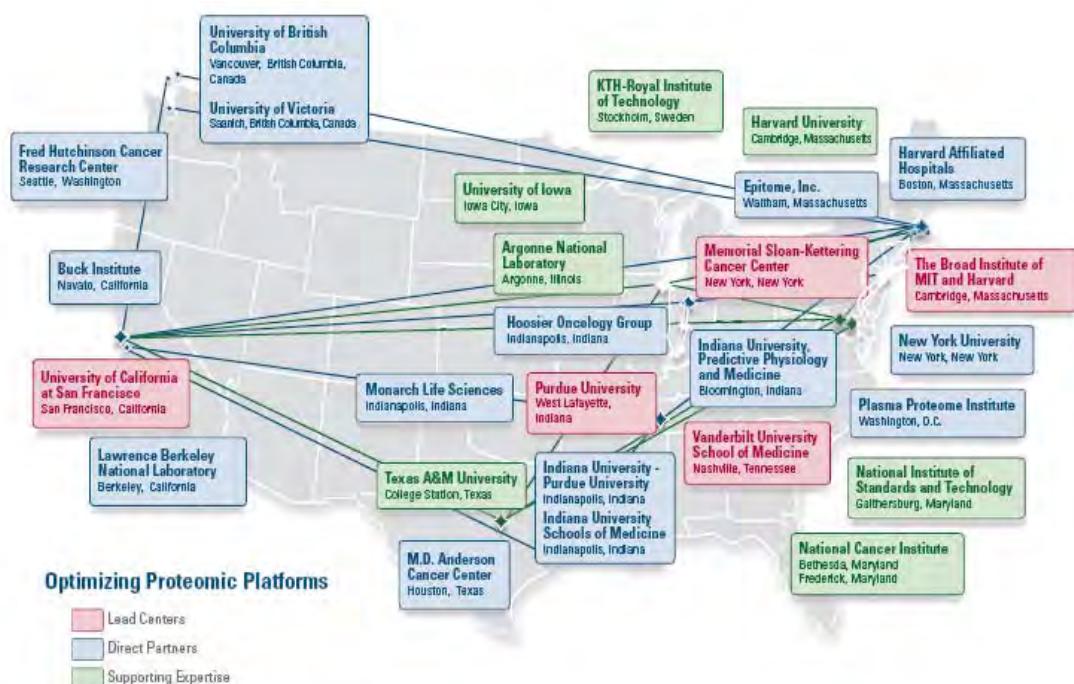
When this initiative began, clinical proteomics had a tarnished image due to some exaggerated claims. It was clear then and now that the technology is the challenge, and CPTAC investigators are developing paradigms to address each part of the analysis process. The ultimate goal of this project will be a pipeline for biomarker discovery in which each step has been rigorously, quantitatively tested. The advances made by CPTAC investigators in the last two years are very encouraging.

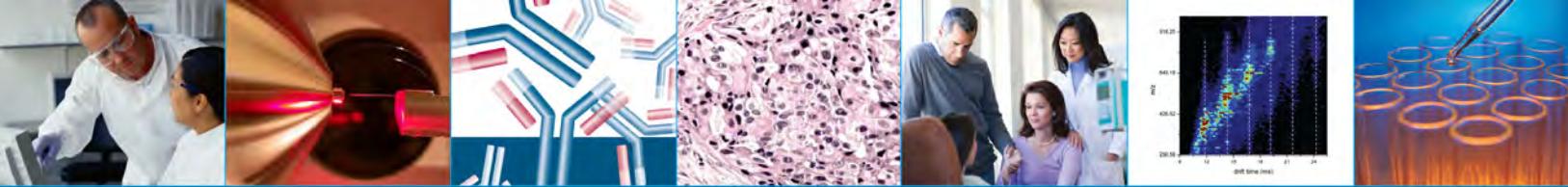
*Leland Hartwell, Ph.D.  
CPTAC Annual Retreat  
November 9, 2008*



Current clinical proteomic research is hampered by a lack of standardized technologies and methodologies, which are critically needed in order to more effectively discover and validate proteins and peptides relevant to cancer, or "biomarkers." To address this critical need, the NCI established a collaborative network of five CPTAC teams in September, 2006. The network extends well beyond these five centers, bringing in expertise from both the public and private sectors to ensure that all the expertise needed is brought together in a single focus (Figure 6). The CPTAC network's ultimate goal is to enable all researchers conducting cancer-related protein research at different laboratories to use proteomic technologies and methodologies to directly compare and analyze their work. This should lead in turn to improved diagnostics, therapies and even prevention of cancer.

**Figure 6. The CPTAC Network**





## CPTAC Teams

### Broad Institute of MIT and Harvard

#### ***“Measuring Cancer Biomarker Candidates by Targeted MS and Antibody Enrichment”***

The Broad Institute Team is developing new targeted, quantitative technologies based on mass spectrometry (MS) that aim to provide the critical bridge between discovery “omics” methods and creation of clinically deployable assays. These methods build on Multiple Reaction Monitoring (MRM) Mass Spectrometry (MS), an approach that has been used previously in clinical chemistry to measure small molecule drugs and metabolites, but now is being tailored for protein biomarkers. For protein biomarker measurements in plasma, they identify one or more “signature” peptides to serve as a quantitative surrogate for each protein to be monitored. Synthetic, stable-isotope labeled versions of each peptide are added into trypsin-digested plasma, and the ratios of labeled to unlabeled (sample-derived) peptides are measured by MRM-MS. These measurements determine the amount of each signature peptide, and hence, the amount of protein from which it is derived. The team has demonstrated that MRM-MS assays can be highly multiplexed enabling the quantitation of 10’s of proteins in a single analysis of patient plasma.

The team is developing a range of sample fractionation and enrichment methods to enable quantitation of proteins at the bottom of the nanogram/mL range or lower in plasma, a range where a large number of clinically useful markers reside. A major focus of the team is in developing and optimizing novel enrichment technology known as stable isotope standards and capture by anti-peptide antibodies (SISCAPA). In this approach, immobilized anti-peptide antibodies are used to capture and enrich the peptides of interest in a single step from blood. The captured peptides are subsequently released and measured by MRM-MS.

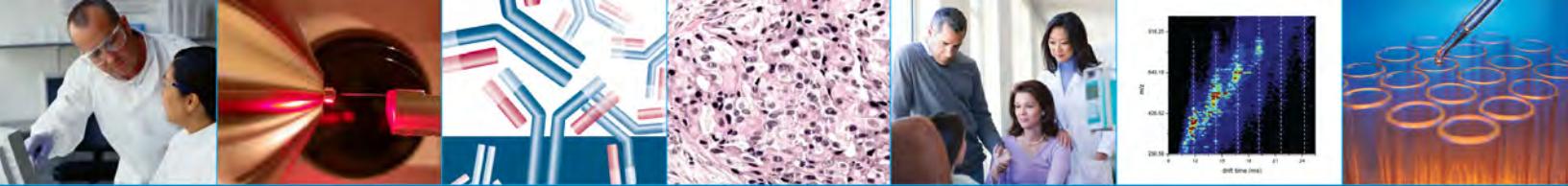
After optimization and characterization, the MRM-MS assays will be used to measure the levels of protein biomarker candidates in plasma from breast cancer patients and compared to the levels in healthy controls. Outcomes will demonstrate 1) the ability to make sensitive/specific assays quickly and inexpensively 2) that assays can be multiplexed, reducing the cost per analyte, and 3) the protocols and technology can be standardized and distributed.

*Principal investigators: Steven Carr, Ph.D., Amanda Paulovich, Ph.D. (Fred Hutchinson Cancer Research Center) and Leigh Anderson, Ph.D. (Plasma Proteome Institute).*

### Memorial Sloan-Kettering Cancer Center

#### ***“Assessment of serum peptide profiling to detect cancer specific patterns”***

The goal of this project is to evaluate and document whether serum peptide patterns, or custom-designed protease assays, have diagnostic value for cancer detection, mark a given clinical outcome, or distinguish clinically insignificant from significant cancer. Such a test could, for example, identify patients with newly diagnosed cancer who might safely avoid surgery or radiation. Investigation of analytical platform robustness and reproducibility is critical to assess the feasibility of future clinical studies.



*Principal Investigator: Paul Tempst, Ph.D.*

**Purdue University**

***"APT: The Analytical Proteomics Team"***

The overall goal of this CPTAC program is to evaluate analytical platforms for validation of breast and prostate cancer biomarker candidates in plasma or serum based on affinity selector targeting of proteins. Proteins and/or peptides thus selected are identified and quantified by either 1) multidimensional mass spectrometry-based methods involving electrospray ionization (ESI) or matrix assisted laser desorption ionization (MALDI), 2) ion mobility separator-based fractionation before multidimensional MS, or 3) immunological arrays on a microfabricated BioCD. Antibody arrays on a BioCD are read either by spinning disc interferometry or laser-induced fluorescence to allow quantification of one to a hundred antigens in one hundred or more plasma samples simultaneously within an hour.

*Principal Investigator: Fred Regnier, Ph.D.*

**University of California, San Francisco**

***"Targeted and global proteomic strategies for early breast cancer detection"***

The overall goal includes the analysis of plasma from breast cancer patients for identification of aberrations in protein splicing and post-translational modifications (glycosylation, phosphorylation, proteolysis and oxidative damage) to profile the cancer proteome and peptidome.

*Principal Investigator: Susan Fisher, Ph.D.*

**Vanderbilt University School of Medicine**

***"Clinical Proteomic Technology Assessment for Cancer"***

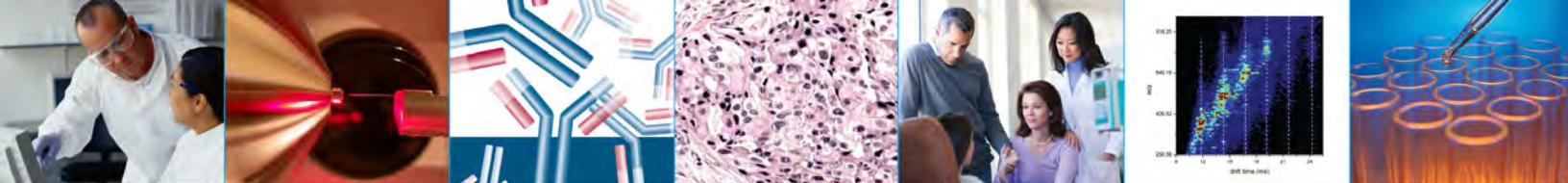
The overall goal is to improve the reproducibility of shotgun and targeted quantitative mass spectrometry as well as to standardize tissue proteomic analysis. Clinical samples from patients with breast, lung and colorectal cancers will be used to define the application of shotgun proteome analysis platforms for the discovery and quantitation of cancer biomarkers.

*Principal Investigator: Daniel Liebler, Ph.D.*

**CPTAC Objectives**

The multidisciplinary CPTAC teams are conducting rigorous assessment of two major technologies currently used to analyze proteins and peptides –MS and affinity capture platforms. Goals include determining if MS and array platforms can be optimized, evaluating fractionation schemes, achieving reproducible protocols, and ultimately enabling quantitative proteomics. Specific objectives include:

- Objective 1:** Evaluating the performance of proteomic technology platforms and standardizing approaches to developing applications of these platforms
- Objective 2:** Establishing systematic ways to standardize proteomic protocols and data analysis among different laboratories



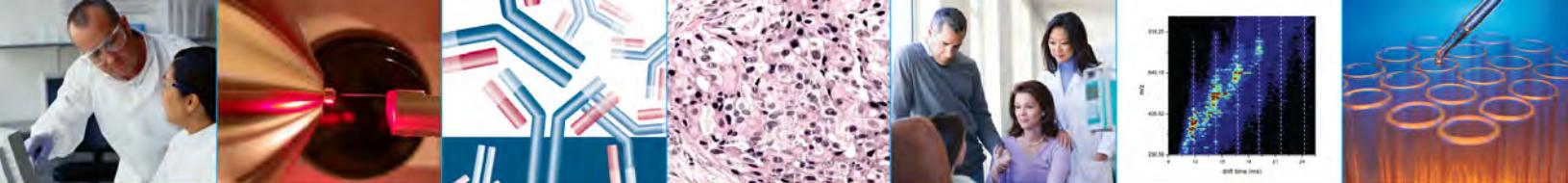
- Objective 3:** Developing and implementing uniform algorithms for sharing bioinformatics and proteomic data and analytical/data mining tools
- Objective 4:** Developing well-characterized material and bioinformatics resources for the entire cancer research community
- Objective 5:** Assessing proteomic platforms for their ability to analyze cancer-relevant proteomic changes in human clinical specimens



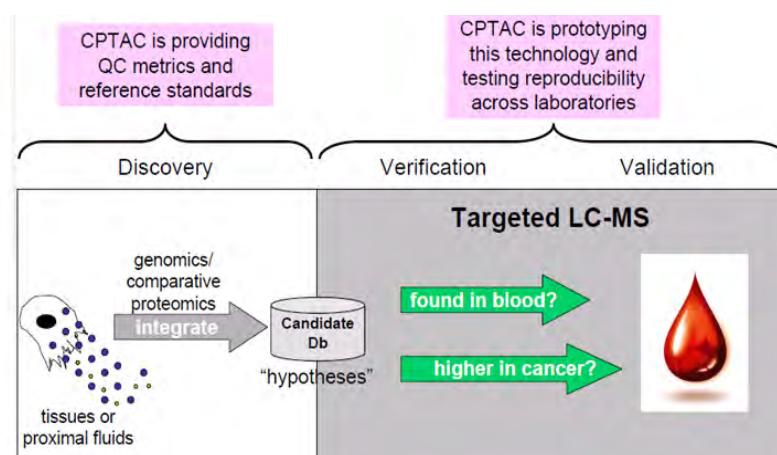
On November 9-11, 2008, CPTAC investigators met in Scottsdale, Arizona, to assess the current status of the CPTAC network relative to its objectives (Figure 7), develop shared, high-level strategies to fulfill pilot goals, and to identify objectives for a possible Phase 2 of the CPTC program.

As discussed in detail below, a number of key milestones have been met during the first two years of the program. Beginning with a critical evaluation of existing technology and methods, the CPTAC teams have designed and carried out true multisite reproducibility studies for two approaches: shotgun unbiased discovery and targeted MRM assays. Standard samples and protocols have been developed and tested by a consortium comprised of the CPTAC teams, NCI and the National Institute of Standards and Technology (NIST). The results, recently presented at scientific and medical, and now submitted for publication, are revealing. As has been expected based on earlier, less well controlled studies (e.g., the Human Proteome Organization [HUPO] plasma proteome exercise), the shotgun approaches produce statistical samples of the peptides in the proteome under study and thus often show significant differences in the sets of peptides from run to run both within and between laboratories (with greater similarity at the protein level). The need for replicate runs to approach asymptotic completeness in proteome coverage is thus an inherent statistical feature of the method. The targeted MRM assays, on the other hand, derived from a widely used accurate quantitation approach for small molecules, can yield results that are accurate, reproducible (in this case across 8 sites), and of wide dynamic range ( $10^3$  to  $10^5$ ), provided attention is restricted to a set of up to several hundred pre-specified peptides. These studies have confirmed the roles and fitness of these two approaches for discovery and verification of candidates in the biomarker pipeline, and provide confidence that both can be practiced effectively in multiple laboratories.

Substantial advances have been made by CPTAC in understanding critical sample requirements, improving the performance of advanced MS instrumentation, and understanding the appropriate structure for a real biomarker pipeline. CPTAC investigators are in the process of creating a conceptual framework for the development of an integrated biomarker development pipeline, a sort of “highway” from discovery to the clinic (Figure 7).

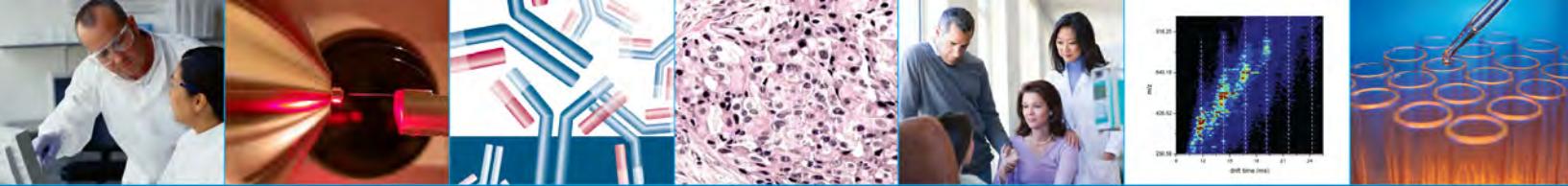


**Figure 7. CPTAC is Building an Integrated Biomarker Development Pipeline**



**At the 2008 CPTAC Annual Retreat, investigators were asked the following question: “What is the single most significant contribution to cancer research that you expect to emerge from the pilot of the CPTAC program?” Some of the responses were as follows:**

- A structured development pipeline for biomarkers applicable to measurement of targets
- Well-characterized reagents and practical, science-based tools made available to the community
- A protein library with abundances that can be used to benchmark future studies
- Restoration of the reputation of clinical proteomics
- Accurate and consistent characterization of proteins using performance metrics for technology platforms used in discovery and validation efforts
- A reproducible process for developing protein or peptide assays that can characterize the lower limits of detection (LOD) in plasma
- A pipeline for generating large numbers of affinity reagents specifically for proteomics

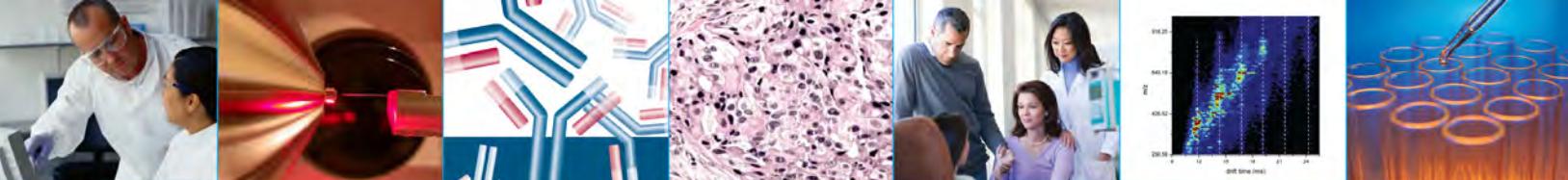


- An expectation instilled in the community that biomarker candidates will be subject to verification, i.e., that they will be assayed in large, meaningful specimen populations with statistical significance to show whether the candidate functions as a biomarker
- Improved predictability of clinical decision-making
- Creation of a critical path for the regulatory process for proteomic assays
- High-quality, well-characterized biospecimens
- A process for enumerating lessons learned and processes to avoid
- Integration of proteomic information with NCI genomics initiatives to narrow the number of candidate biomarkers
- Well-defined parameters for adoption by the community
- Standardization of specimen protocols to support new standards of care and effect a reimbursable expense for pathologists
- An identified biomarker that is well-characterized and clinically useful

### 1.2.2 Advanced Platforms and Computational Sciences Awards

The Advanced Proteomic Platforms and Computational Sciences initiative is a comprehensive program focused on the development of innovative new tools, reagents, and the enabling of technologies for protein/peptide measurement, such as algorithm development and computational methods to interrogate emerging pre-processed data sets. It also sets out to establish the Advanced Platforms, Data Analysis Methods, and Computational Sciences components of the NCI Clinical Proteomic Technologies for Cancer. The Advanced Proteomic Platforms and Computational Sciences initiative supports two focus areas for protein measurement technology and application in cancer research:

- Development of innovative high-throughput technology for protein and peptide detection, recognition, measurement, and characterization in biological fluids that will overcome current barriers in protein/peptide feature detection, identification, quantification, and validation.
- Development of computational, statistical, and mathematical approaches for the analysis, processing, and facile exchange of large proteomic data sets.



Advancing the technological and analytical capabilities in proteomic research will allow the research community to better characterize and understand the differences between the normal and diseased human proteome and to develop diagnostic and treatment procedures based on these distinctions.

## Advanced Proteomic Platforms

### **Proteomic Phosphopeptide Chip Technology for Protein Profiling**

Principal Investigator: Xiaolin Gao, Ph.D., University of Houston

### **Global Production of Disease-Specific Monoclonal Antibodies**

Principal Investigator: Barry L. Karger, Ph.D., Northeastern University

### **Top-Down Mass Spectrometry of Salivary Fluids for Cancer Assessment**

Principal Investigator: Joseph A. Loo, Ph.D., University of California Los Angeles

### **A New Platform to Screen Serum for Cancer Membrane Proteins**

Principal Investigator: Daniel B. Martin, M.D., Institute for Systems Biology

### **A Proteomics Approach to Ubiquitination**

Principal Investigator: Junmin Peng, Ph.D., Emory University

### **A Proteomics Platform for Quantitative, Ultra-High Throughput, and Ultra-Sensitive Biomarker Discovery**

Principal Investigator: Richard D. Smith, Ph.D., Battelle Pacific Northwest Laboratories

### **Aptamer-Based Proteomic Analysis for Cancer Signatures**

Principal Investigator: Stephen P. Walton, Ph.D., Michigan State University

## Computational Sciences

### **Proteomic Characterization of Alternate Splicing and cSNP Protein Isoforms**

Principal Investigator: Nathan J. Edwards, Ph.D., Georgetown University

### **Enhancement of MS Signal Processing Toward Improved Cancer Biomarker Discovery**

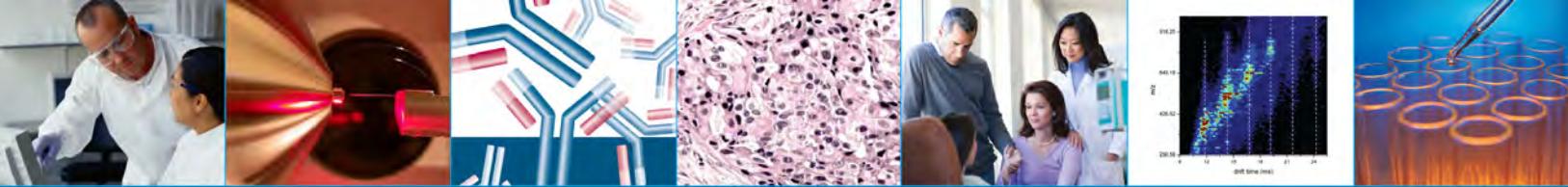
Principal Investigator: Dariya Malyarenko, Ph.D., College of William and Mary

### **A Platform for Pattern-Based Proteomic Biomarker Discovery**

Principal Investigator: Denkanikota Mani, Ph.D., Massachusetts Institute of Technology

### **Analysis and Statistical Validation of Proteomic Datasets**

Principal Investigator: Alexey I. Nesvizhskii, Ph.D., University of Michigan



### **Quantitative Methods for Spectral and Image Data in Proteomics Research**

Principal Investigator: Timothy W. Randolph, Ph.D., Fred Hutchinson Cancer Research Center

### **Computational Tools for Cancer Proteomics**

Principal Investigator: Katheryn A. Resing, Ph.D., University of Colorado at Boulder

### **New Proteomic Algorithms to Identify Mutant or Modified Proteins**

Principal Investigator: David L. Tabb, Ph.D., Vanderbilt University

### **PICquant-An Integrated Platform for Biomarker Discovery**

Principal Investigator: Dennis J. Templeton, Ph.D., University of Virginia

Full biographies of the Advanced Proteomic Platforms and Computational Sciences investigators can be found in Appendix 4.2.

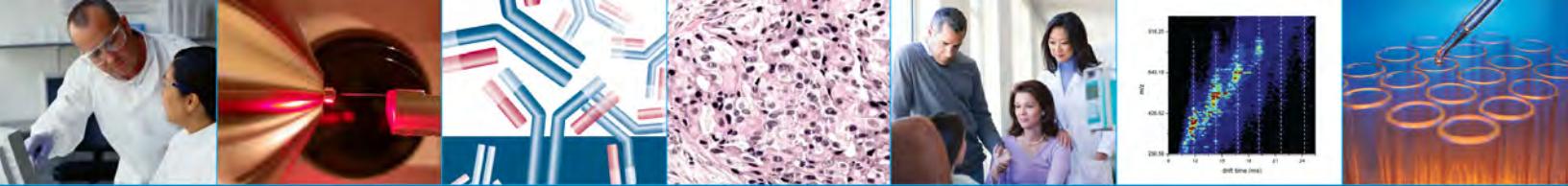
#### **1.2.3 The Proteomic Reagents and Resources Core**

In order for the next phase of biomedical research to occur (incorporating proteomics with genomics), high-quality proteins and well-characterized/validated affinity capture reagents (e.g. antibodies) comprising the human proteome are needed for accessibility across the scientific community. A reality check on today's antibody market is that while tens of thousands of reagents are commercially available, few are well-characterized with many being highly variable in terms of quality. Furthermore, few of these reagents are directed against novel (investigator-driven) targets, due to intellectual property issues. The lack of such high-quality reagents is a significant bottleneck for clinical proteomics. This barrier was recognized by the NCI, which led to the development of the Reagents & Resources Core of its CPTC program.

Outlining a strategic plan to address this bottleneck, the NCI held a workshop in December, 2005. Attendees were from the public, private, academic, and international institutions. Key considerations outlined by the scientific community included:

- Making antigens freely available
- Supporting renewable antibody production (e.g. monoclonal antibodies)
- Enabling antibody characterization using Standard Operating Procedure (SOP)-driven protocols
- Supporting distribution of antibodies and hybridomas with no intellectual property in order to promote multiplex affinity capture platform development

It is these recommendations that NCI's CPTC Reagents and Resources Core is based upon. This program acts as a catalyst to spur the development of pivotal resources



(such as antibodies) that serves the entire research community, including both public and private entities. These resources are necessary to accelerate biomarker discovery and validation, translational research, molecular diagnostics, and therapeutic monitoring.

Reagents are available through a central portal (<http://antibodies.cancer.gov>) that is accessible through the CPTC website.

### 1.3 Governance

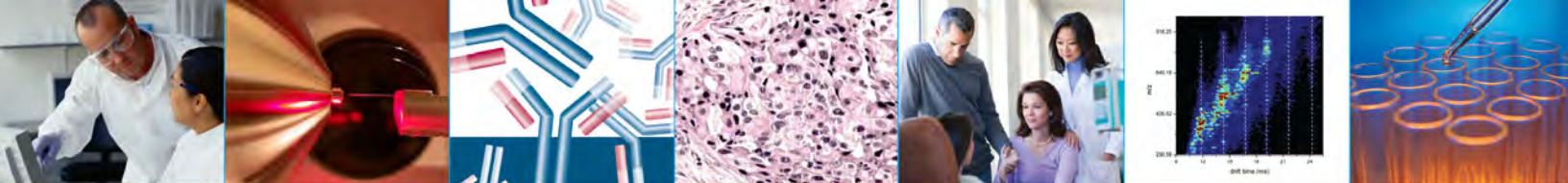
#### 1.3.1 Program Coordinating Committee

A Program Coordinating Committee (PCC) was established as a CPTAC governing body that sets policy and approves procedures that all participating teams have to follow. In addition, PCC oversees the integration of the CPTAC teams as part of the entire CPTC. Voting members of the PCC include the Principal Investigator (PI) or a PI-designated senior scientist from each CPTAC team and one person representing NCI Program staff (the NCI Project Coordinator). Additional non-voting members with specific scientific expertise in disciplines not readily provided by the PCC members have been added to ensure an adequate review of the research infrastructure.

On a semi-annual basis, at minimum, the PCC assesses the progress made by the participating CPTAC teams, makes recommendations regarding strategic decisions and program goals, and provides guidelines regarding experimental design, sharing of protocols and reagents, and the allocation of resources. The PIs are expected to accept and implement the recommendations.

Other activities of the PCC include:

- Prioritization of materials/reagents for further characterization and standardization by the participating CPTAC teams;
- Evaluation of proteomic standard reference materials, standard protocols, and bioinformatics tools/standards developed or adopted by the individual CPTAC teams and recommendations for incorporating respective resources/procedures into the research efforts of all CPTAC teams and broader scientific community;
- Monitoring of concurrent developments in clinical proteomics external to the CPTAC team network;
- Facilitating communication and data sharing among the CPTAC teams;
- Coordinating the compatibility of generated data and bioinformatics tools among participating CPTAC teams with caBIG™;
- Establishing an interactive web site providing both open public access and password-protected access for CPTAC team participants;
- Organizing conference calls and meetings for participating investigators; and



- Coordinating efforts with the NCI to ensure that resources, technologies, and knowledge gained from the program are made publicly available and widely disseminated throughout the cancer research community.

Responsibilities of CPTAC members are provided in Appendix 4.3.

### PCC Members

Steve Carr, Ph.D., Broad Institute of MIT and Harvard

Susan Fisher, Ph.D., University of California, San Francisco

Dan Liebler, Ph.D., Vanderbilt University

Paul Tempst, Ph.D., Memorial Sloan-Kettering Cancer Center

Fred Regnier, Ph.D., Purdue University

Henry Rodriguez, Ph.D., M.B.A., National Cancer Institute

### Ad-hoc PCC members

Leigh Anderson, Ph.D., Plasma Proteome Institute

Bradford W. Gibson, Ph.D., Buck Institute of Age Research

Joe Gray, Ph.D., Lawrence Berkeley National Laboratory

Lee Hartwell, Ph.D., Fred Hutchinson Cancer Research Center

Gordon Mills, M.D., Ph.D., M.D. Anderson Cancer Center

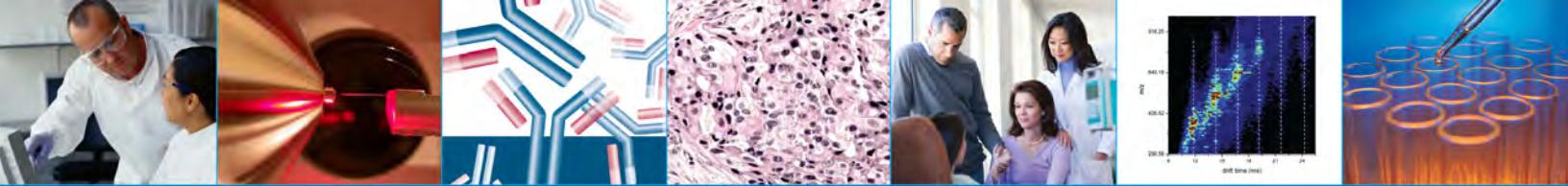
Amanda G. Paulovich, M.D., Ph.D., Fred Hutchinson Cancer Research Center

David Ransohoff, M.D., University of North Carolina Lineberger Comprehensive Cancer Center

Steven J. Skates, Ph.D., Massachusetts General Hospital

Full biographies of the PCC committee members can be found in Appendix 4.4.

### **CPTAC Internal Advisory Committee**



Each CPTAC team has an Internal Advisory Committee (IAC). The IAC consists of individuals from within the CPTAC team as well as other components of the applicant institutions that have been integral partners in developing the experimental design plans for the CPTAC team.

The IAC supports the Team Leader in coordinating and monitoring the activities of the CPTAC teams. The proposed IAC defines the chain of responsibility within the CPTAC team for decision-making and administration beginning at the level of the Team Leader and including all key staff. The IAC works with the Team Leader to establish a timeline for proposed activities and to develop opportunities for information exchange, seminar presentations, and research training opportunities.

Official communication by each respective IAC is represented through each respective Team Leader at the PCC.

### ***CPTAC Working Groups***

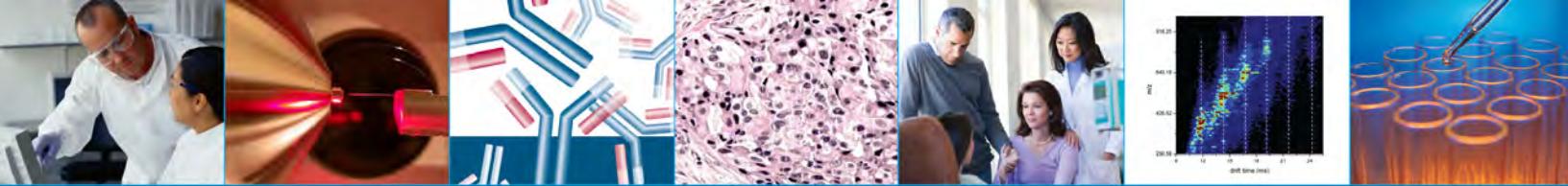
The CPTAC PCC is the governing body for the CPTAC project. The working groups are organized as subcommittees of the PCC. The membership of each group is composed of one or more subject matter experts designated by the Team Leader, from each of the funded CPTAC components. In making such appointments, considerations must be based on technical expertise.

Each working group has a Chair. Working group members must be open to the views and concerns of all interests as they bear on the working groups' work product. It is the responsibility of the PCC to ensure that all views are considered in the final work product. Working groups should not be formed to address broad issues involving policy choices.

Each working group independently develops recommendations according to their specific mission. These recommendations are then presented to the PCC for approval and subsequent implementation.

The initial working group discussions are to cover a subset of topics listed below and then as needed, new ones are to be added and obsolete ones terminated. The working groups will minimally meet for monthly teleconferences. Additionally, Chairs of all the working groups are to hold a teleconference post working group teleconferences, and pre-PCC monthly teleconferences.

Current working groups are: Unbiased Discovery Studies; Verification Studies; Post-Translational Modification Needs and Production; Biospecimens; Analyte Selection; Protein Standards; Plasma Sample; Digestion; Data Analysis, Storage and Dissemination; Cell Line Selection and Lysate; Yeast Lysate. Working Groups are added or dissolved as deemed necessary by the PCC.



### 1.3.2 Communications Plan

Within the network, the various committees and working groups have a rigorous communication structure that ensures innovation and creativity while maintaining focus on the program's overall mission. Innovative ideas bubble up from CPTAC centers or within working groups and are shared with the PCC. The PCC then aligns each idea with the mission of the program and oversees the execution of each inter-laboratory project.

Working groups have a monthly conference call in the first two weeks of the month. During the third week of the month, the working group chairs report to the chair of the PCC via teleconference. In the fourth week of the month, the PCC has a call to discuss scientific direction and business items. Additional emails, phone calls, and correspondence are thereby anchored in a foundation of regular communication throughout the program.

## 1.4 Program Management

### 1.4.1 Scientific Working Groups

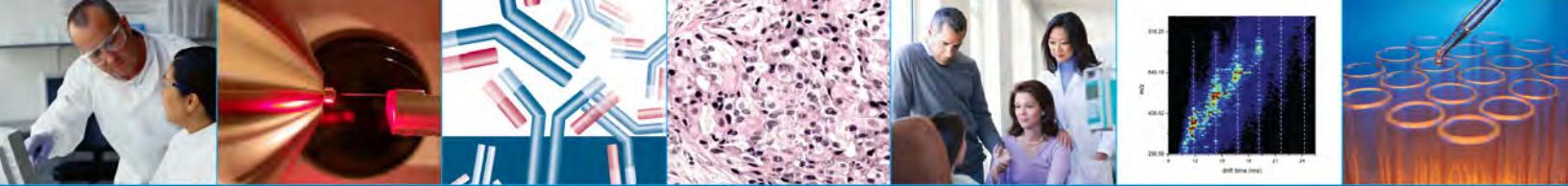
The approach for all CPTAC collaborative studies is the comparative analysis of data generated across participating labs. The development of a working infrastructure is a requirement for the success of the program and therefore a multitude of Working Groups were created to meet this goal. Working Groups can be eliminated or created as CPTAC endeavors evolve and the need for new experiments and protocols arise.

The CPTAC Working Group Membership Lists can be found in Appendix 4.5.

### 1.4.2 Site Visits

The purpose of site visits is to enable CPTC to evaluate an organization that has been awarded a cooperative agreement, or contract with stated deliverables. Visits also allow time for Team Leaders, investigators and CPTC management to discuss and resolve issues associated with an awarded project. Additionally, site visits provide an opportunity to get to know CPTAC members (including junior staff, post-docs, and graduate students) and to become familiar with research facilities. Overall, site visits are designed for CPTC to help the CPTAC center successfully meet the objectives of the program.

Site visit Guidelines and Checklist can be found in Appendix 4.6.



### 1.4.3. Informatics and caBIG™ Connectivity

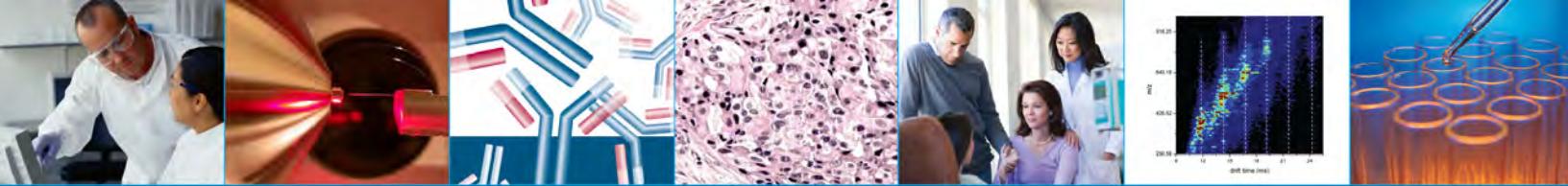
CPTC supports a number of computational efforts that stand to revolutionize proteomic data analysis, storage, sharing, and provenance. These projects aim to develop computational, statistical, and mathematical approaches for the analysis, processing, and facile exchange of large proteomic data sets. CPTC is advancing these areas with the added goal of making proteomics data transparent and accessible to protein scientists as well as other members of the biomedical community.



In June 2008, members of CPTC met with representatives from CBIIT to strategize on how to further capitalize on the opportunities to advance transparency and accessibility of proteomics data. Out of that workshop came a strategic plan for caBIG™ to adopt the Tranche repository, which entailed development of a controlled vocabulary that includes terms required to describe a proteomics experiment. Successful completion of this project will open up proteomics data to all databases and software tools on caBIG™.

At the most fundamental level of data, the Tranche Project, developed by Phil Andrews of the University of Michigan, is a free and open source file sharing tool that enables collections of computers to easily share and cite scientific data sets. Designed and built with scientists and researchers in mind, Tranche essentially solves the data sharing problem in a secure and scalable fashion. Tranche uses secure distributed file sharing network concepts mixed with modern encryption to make a secure distributed file system that is well-suited for any size data and independent of any particular centralized authority. With nearly 5 TB of MS raw files, Tranche is the primary repository for proteomics data. Tranche has served as the repository for the CPTAC network, hosting all inter-laboratory data and metadata. In 2009, Tranche and its associated annotation tool will become caBIG™-compliant, making CPTC data accessible to the broader biomedical research community.

Another caBIG™-compliant data repository, Computational Proteomics Analysis System (CPAS), tracks and analyzes proteomics data throughout an experiment. CPAS incorporates a robust data pipeline for importing and processing MS/MS data from raw and mzXML data files. The pipeline manages the chain of processing steps needed to infer protein identifications and expression levels from the output of a mass spectrometer. The pipeline integrates with leading search engines including X!Tandem™, Mascot™, and Sequest™. The pipeline also integrates PeptideProphet™, ProteinProphet™, and XPRESS™, all components of the Trans Proteomic Pipeline™ from the Institute for Systems Biology. For analysis, CPAS displays the search results



from one or many runs in a web browser, enabling one to filter, sort, customize, compare, and export experimental runs. Data can be shared securely with collaborators inside or outside of an organization while exercising fine-grained control over permissions.

Today, CPAS powers proteomics repositories at the Fred Hutchinson Cancer Research Center, Cedars-Sinai Medical Center, the National Institutes of Health, and many others, where it is central to efforts to identify predictive biomarkers for cancer.

#### 1.4.4 Supporting Partnerships



##### ***Standards Development and Data Analysis***

The National Institute of Standards and Technology (NIST) has expertise in the qualification and characterization of peptide and protein based standards. NCI has entered into an Interagency Agreement with NIST to develop mass spectrometry (MS) assessment materials to be used by the CPTAC teams with the intention of developing proteomic standard reference materials (SRM). These SRMs, designed to assess the performance metrics of various instruments, will be the first of their kind developed by the NCI and will help to evaluate and compare existing proteomic technologies and compare these with emerging technologies of interest to the clinical cancer community. SRMs will be provided to the scientific community through the CPTC Reagents and Resources Core component of the program.

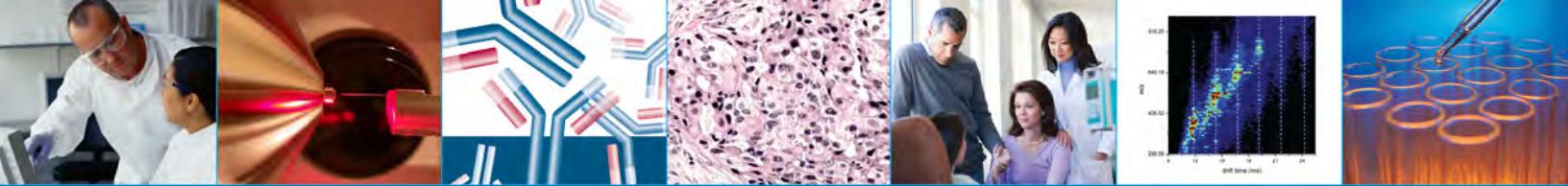


##### ***Experimental Design and Statistical Analysis Expertise***

NCI has partnered with researchers from Texas A&M University and the National Institute of Statistical Sciences (NISS) to develop a core competency in the technology program that supports metrology, study design, statistical analysis, and methodological approaches to applying proteomic technology platforms toward clinical measurement.

#### 1.4.5 Software Commercialization

CPTC supports a number of software development projects that stand to revolutionize proteomic data analysis, sharing, storage, and provenance. One such project, CPAS, is now fully compliant on caGRID, the federated, interoperable network of caBIG™. Recognizing the value of CPAS, Genologics, a leading provider of laboratory information management software for proteomics, has integrated CPAS into its Proteus software suite.



**FRED HUTCHINSON CANCER RESEARCH CENTER**  
A LIFE OF SCIENCE

**Data Repository - CPAS**

CPAS and Proteus™ integrate seamlessly to provide a robust solution for cancer researchers. Proteus and CPAS (Computational Portal and Analysis System) operate in tandem to provide an integrated lab and data management system and proteomics data repository that allows cancer researchers from separate labs around the world to analyze, store, share, and publish data from proteomics experiments using mass spectrometry. CPAS is an extensible, web-based science portal that acts as a repository and a collaboration tool to share results. Proteus is an easy to use, configurable, integrated lab and scientific data management system for proteomics research.

**LabKey SOFTWARE**

GenoLogics and LabKey Software (along with the Fred Hutchinson Cancer Research Center) have teamed to combine the two offerings into a configurable, collaborative, end-to-end cancer research solution. In the combined solution, Proteus manages experimental data and tracks laboratory workflows up to and including the generation of mass spectrometry data output files. Proteus then exports the mass spec data along with a description of the sample processing steps to CPAS using the XAR (Experiment ARchive) format. CPAS runs peptide and protein search algorithms, stores the results, and presents a viewer that helps researchers to understand results and share them. This close integration is possible because both CPAS and Proteus embrace standard data formats and open standards software tools for proteomics experiments.

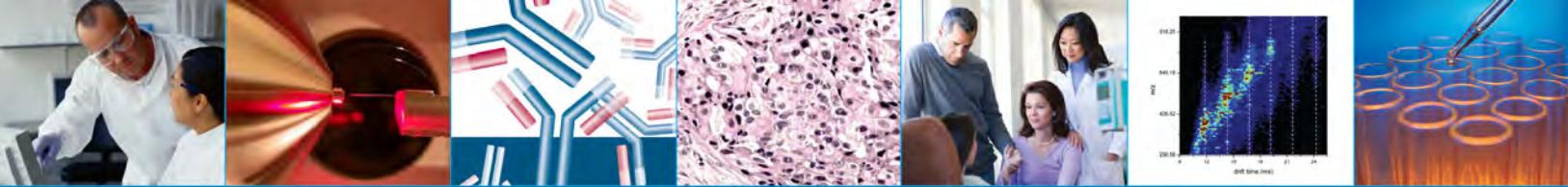
ENABLING TEAM SCIENCE

[www.genologics.com](http://www.genologics.com)

In addition, several of our sites and outside the CPTC Network received a "Bioinformatics Platform Dissemination" Award from the Canary Foundation, to adopt our CPAS- now offered through CBIIT.

**Carl Schaefer, Ph.D.**  
*Director, Cancer Genome Anatomy Project*  
**Center for Bioinformatics and Information Technology**  
**National Cancer Institute**

The Computational Proteomics Analysis System (CPAS) originated as a laboratory information management system for proteomics. Early versions provided solid support for recording experimental protocols and for running peptide identification pipelines. However, CPAS was not originally intended or designed to support queries across experiments, protocols, or runs. More recent versions do allow users to compare lists of

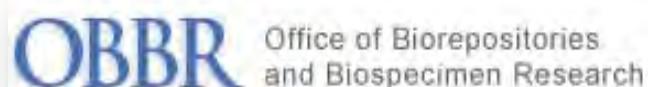


peptides across runs, but user query capabilities remain limited. NCI recognized that some users might wish to execute warehouse-style queries, such as finding all runs in which certain proteins were identified or finding all experiments that used a given protocol. Implementing queries such as these could entail significant amounts of software development. After considering the options, NCI decided to use BioMart to build the desired capability.

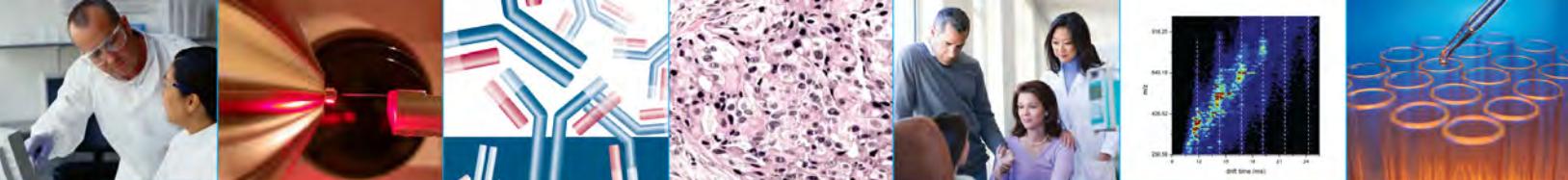
Developed jointly by the Ontario Institute for Cancer Research and the European Bioinformatics Institute, BioMart (<http://www.biomart.org/>) is a software system for building data warehouses and servicing user queries. In the case of CPAS, NCI determined that the simplest approach was to export data from the standard CPAS database to a parallel database that conforms to the BioMart model. The code to perform the export and import operations is automatically generated from specifications of the exporting and importing databases. Likewise, the entire user query interface is automatically generated from a specification of how database tables can connect to each other by common data fields. In this way, the BioMart interface to CPAS was created without having to design, code, and test custom software for the user interface and business logic. The BioMart interface is running on NCI's installation of CPAS at <http://cpas.nci.nih.gov>.

#### 1.4.6 Biospecimens and NCI Best Practices

Access to high quality biospecimens presents a major challenge within clinical proteomics. Since its inception, CPTC has worked closely with the Office of Biorepositories and Biospecimen Research (OBBR) to ensure quality, legality, and assurance of biospecimens.



At the launch of CPTAC, the five centers quickly realized that each center was collecting plasma associated with women who may have breast cancer. Recognizing the opportunity to create a valuable collection of specimens, Dr. Steven J. Skates (Dana-Farber/Harvard Cancer Center) and Dr. David F. Ransohoff (University of North Carolina at Chapel Hill) organized the investigators responsible for sample collection at each core center into a Biospecimen Working Group. This Working Group held a strategy meeting in June of 2007 to adopt a common SOP for biospecimen cohort, collection, processing, and storage. A portion of each plasma sample is sent to a central biorepository in NCI-Frederick. Using a central biorepository, each center gains access to a much larger number of patient samples, and greatly increases the statistical power available for experiments. Members of OBBR, especially Dr. Helen Moore, have actively



participated in this process, using it as a case study for biospecimen collection within multi-center, multi-disciplinary projects.

In addition, NCI's Clinical Proteomic Technologies for Cancer program staff serve as proteomics scientific technical experts to OBBR's Biospecimen Research Network.

#### 1.4.7 Communication and Working with Outside Organizations (Domestic and International)

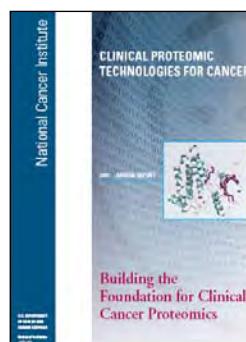
##### 1.4.7.1 Communication Materials

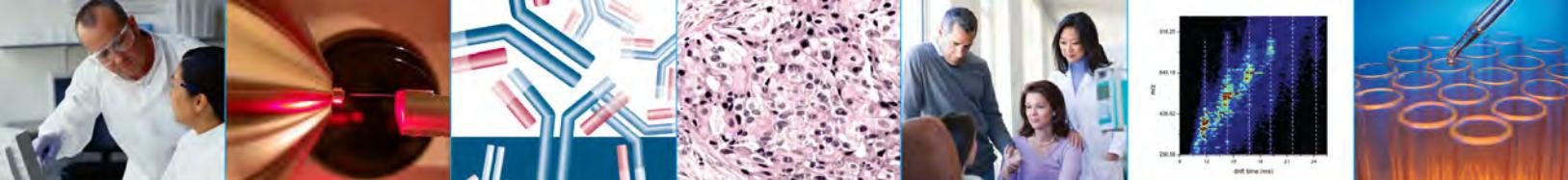
A variety of materials have been developed to aid in communicating CPTC's mission and milestones to a number of audiences. These have all been made available through the CPTC website (<http://proteomics.cancer.gov/>), the center of all communications in today's digital world.



#### ***Annual Reports***

The CPTC Annual Reports chronicle the five-year program, offering highlights and progress reports from each of the CPTC components, as well as information about the collaborative community and public/private partnerships that are a large part of CPTC.





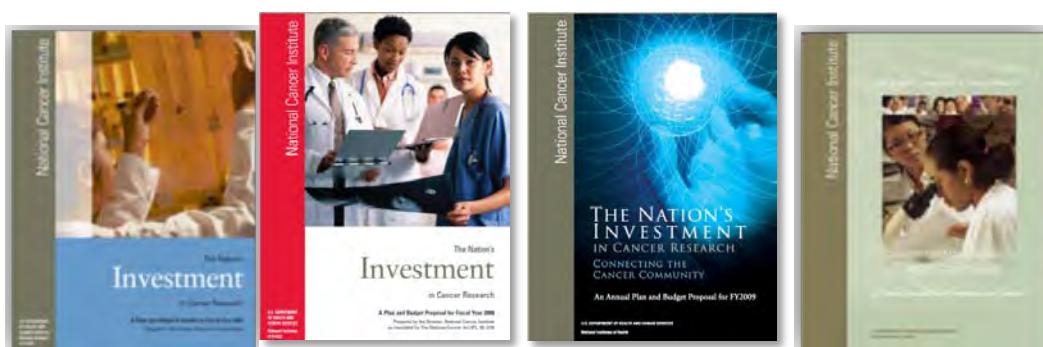
## NCI Annual Plans and Budget Proposals

**FY2007:** We need to refine the technology so that it can find “a needle in the haystack” with unprecedented reliability.

**FY2008:** CPTC developing the standards needed to characterize protein markers for very early detection of cancer.

**FY2009:** Cancer prevention is defined by use of advanced tools and technologies - such as those employed in proteomics.

**FY2010:** CPTC’s Antibody Characterization Program: Well-characterized antibodies will be vital to advancing molecular diagnostic techniques like proteomics.

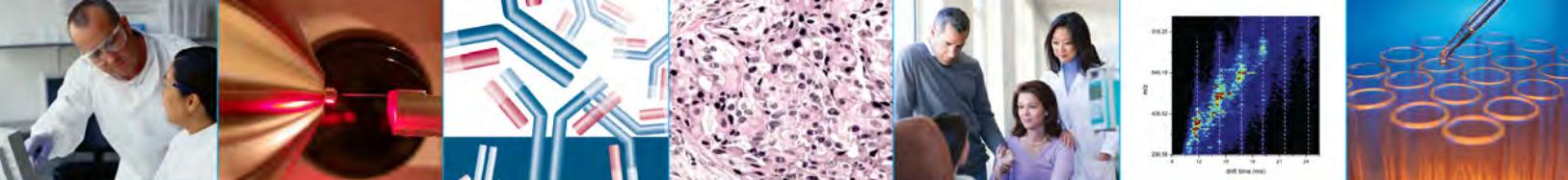


## eProtein newsletter

The eProtein newsletter is a new communication tool that was launched in December 2008 as a way to proactively reach out to the community on a quarterly basis.



The eProtein newsletters can be found in Appendix 4.7.



## CPTC Program Milestones Timeline

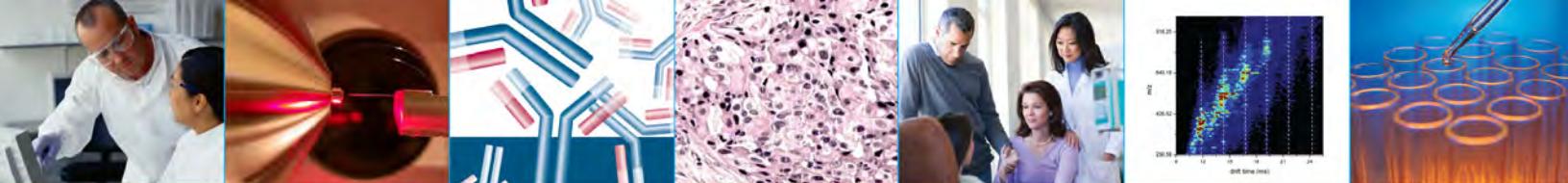
An interactive CPTC Program Milestones timeline was created to help all audiences “see” the accomplishments being made by the initiative even if they don’t fully understand the complex technical challenges that are being overcome. This was launched in December, 2008.



## Proteomics Primer

The proteomics primer is an educational tool that was created as a way to educate lay audiences on the basics of proteomics and its promise for the early detection and treatment of cancer.





### Pioneers of Proteomics

The *Pioneers of Proteomics* video series presents leaders in the field of proteomics discussing recent developments in the field, the inherent challenges in studying proteins and significant opportunities afforded by this rapidly evolving field. The series includes:

- Leland Hartwell, Ph.D.
- Joshua LaBaer, M.D., Ph.D.
- Ruedi Aebersold, Ph.D.
- Stanley Hefta, Ph.D.
- John Yates, Ph.D.
- Gilbert Omenn, M.D., Ph.D.
- Richard Caprioli, Ph.D.
- Catherine Fenselau, Ph.D.
- Mark Boguski, M.D., Ph.D.

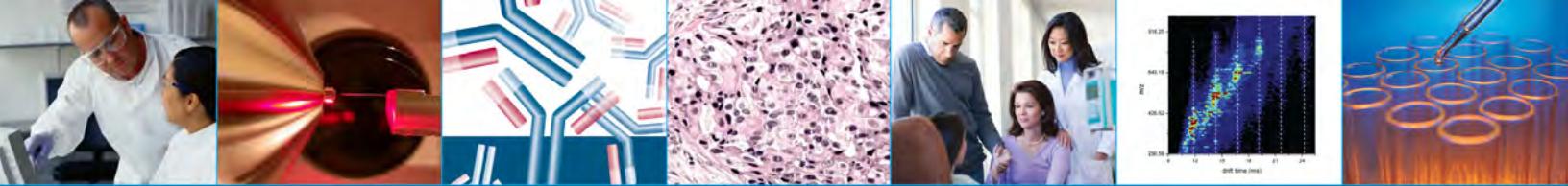
### Video Tutorial

An interactive video tutorial entitled, “Explore Proteomic Technologies and Cancer” was developed shortly after the program was launched that walks viewers through the process for discovering clinical biomarkers, from patients to samples and back to patients.



### 1.4.7.2 Advocacy Outreach Activities

CPTC places a premium on communicating with research stakeholders to ensure that opportunities to incorporate unique perspectives are explored fully and optimally. This commitment to communication and outreach does not stop with investigators, clinicians,



and private sector representatives who translate discoveries from the bench to the bedside.

Diagnostics and therapeutics developed using support from CPTC research programs ultimately benefit patients and those disease-free individuals who will benefit from preventive approaches. CPTC engages with these beneficiaries through an organized outreach program that includes direct involvement with and input from representatives of the advocacy community, including NCI's Consumer Advocates in Research and Related Activities (CARRA). CARRA members participate in a wide range of NCI activities and represent the collective viewpoint of people affected by cancer. CARRA members participate in a variety of NCI activities involving scientific research and communication of scientific results, including but not limited to, sitting on committees and boards, and attending meetings, workshops, and site visits.

The partnership between CPTC and CARRA is a truly symbiotic relationship. Patients and advocates are informed about research in proteomics-based technologies with the potential to improve detection and treatment of cancer, and this newly informed group ultimately helps accelerate the adoption of these technologies as they become clinically available. Meanwhile, researchers and clinicians have a constant reminder of the patients that their work will impact. As this information sharing continues, the collaboration between CPTC and CARRA strengthens, and a better outcome for patients comes within reach.

**Elizabeth Nielson**  
**CARRA Program Manager**  
**National Cancer Institute**

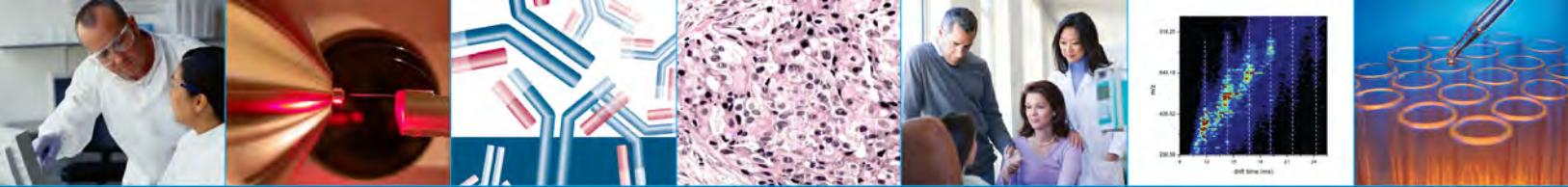
As Director of the CPTC initiative, Dr. Henry Rodriguez values the involvement of cancer



Consumer Advocates in Research and Related Activities  
A Program of NCI's Office of Advocacy Relations

advocates to bring the unique patient perspective to his board, meetings and publications. In fact, when CPTC was first launched in 2006, Consumer Advocates in Research and Related Activities (CARRA) members were directly involved in helping to develop the initiative's mission. The involvement of CARRA members in this work is a helpful reminder of the end goal, which is to discover and treat cancer as early as possible in patients.

Explaining clinical proteomics—and thus the purpose of CPTC—to the non-scientific community is an extremely daunting task. To help communicate complex scientific messages to patients and advocates, CARRA members worked closely with CPTC to create a brochure entitled, "Clinical Cancer Proteomics: What it Means and What it Means for You."



The brochure was developed for two purposes. First, CPTC wanted to recruit CARRA members to serve on the CPTC board to share the needs of those affected by cancer. The brochure would give CARRA members the background needed to be effective members of the board and know how to critically review research. Second, CPTC wanted to promote clinical proteomics because in terms of advanced technologies ultimately the clinical community needs to adopt them and the public needs to understand them. CARRA members can help create a bridge because they represent a community of patients who are educated on the technologies.

Thanks to the direction, focus and commitment from CARRA members recruited by Dr. Rodriguez, a five-page brochure is now available to the public, outlining in very simple terms the importance of clinical cancer proteomics and the work of CPTC. Today, with brochure in hand, Dr. Rodriguez and team can effectively and efficiently explain their goal to the masses.



Brochure available at <http://proteomics.cancer.gov/library/brochure.asp>

### **Podcast Series**

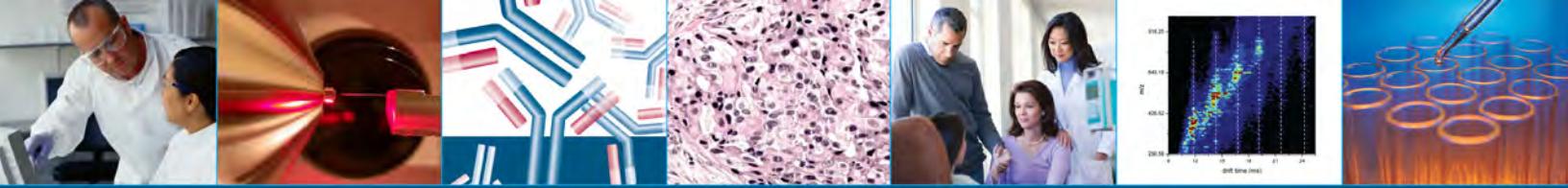
**“The Promise of Proteomics for Personalized Medicine”** is the first of a three part podcast series developed to provide listeners with information about research advances being made in proteomics. This podcast can be found at:  
<http://proteomincs.cancer.gov/library/podcasts.asp>

### **Promise and Reality of Proteomics Webinar**

The NCI Office of Advocacy Relations hosted a webinar on March 19, 2009, on the role of protein science in the early detection of cancer as part of the Understanding NCI: Toll-Free Teleconference Series. Dr. Rodriguez and other speakers discussed the challenges facing clinical proteomics and the innovative ways that NCI's CPTC initiative aims to develop new, more refined, efficient, and reliable biomarkers discovery and verification pipelines. This webinar can be found at:  
<http://proteomincs.cancer.gov/library/webinars.asp>

Speakers included:

*Henry Rodriguez, Ph.D., M.B.A.*, Director, Clinical Proteomic Technologies for Cancer, NCI Center of Strategic Scientific Initiatives



*Amanda G. Paulovich, M.D., Ph.D., Oncologist and Cancer Geneticist, Director, Early Detection Initiative, Fred Hutchinson Cancer Research Center*  
*Elda Railey, Co-Founder of Research Advocacy Network*

**Elda Railey**  
**Co-Founder, Research Advocacy Network**  
**Webinar Response**  
**The Promise and the Reality of Proteomics**

## Research Advocacy Network

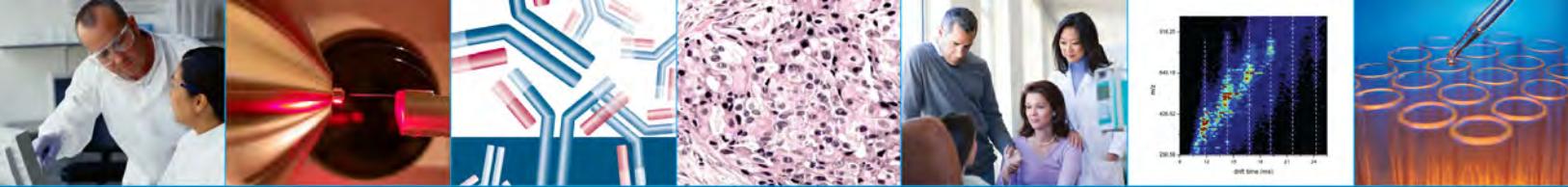
For years, we have been promised more personalized medicine and targeted therapies for cancer and today we have learned more of the real promise for cancer detection and treatment through the study of proteomics. There is much promise through proteomics but the reality is that the gold standard of detection in most cancers currently lies with imaging. The issue of cost savings through early detection methods is very important as the strain on our healthcare system is widely felt as well. However, the cost savings of early detection methods are negated when the results are not reliable or the costs of the tests outweigh their validity.

Part of the “omics” promise is the development of better candidates for drug therapies and for early cancer diagnostics where the cure rate is improved and the cost of treatment can be reduced both in human suffering and dollars. This promise is especially enticing in harder to diagnose cancers such as squamous cell head and neck cancer.

I must admit that it was still confusing to me as I was preparing for this response what the difference is between genetics/genomics and proteomics. After studying the materials from NCI, I now understand better that genes give a glimpse of what MAY occur and proteomics can help understand what is happening in REAL TIME. The reality for a patient is that it is not important what type of technology or “omics” science results in the best detection methods and personalized treatment choices, but it is very important to patients that the results returned by these technologies are accurate and reliable.

When donating biospecimens for research it matters that “that piece of me,” whether it is blood, serum, or other biospecimens, is used to gain the maximum amount of information and contribute to the knowledge base to fight cancer. We also want to be assured that our privacy is protected.

It is important to get the word out... I had a bit of a reality check myself when I updated my Facebook status about my preparation for this teleconference today. I had several ask ... can you give me the Proteomics for Dummies? so they could understand what I was doing. I must admit that my response was difficult and that even though I have been in cancer advocacy for many years it was not easy to make this technology a concept



relevant to our everyday lives. Yet this is where much of the research investment dollars and the state of science is headed. It was recently published that colon cancer patients who knew about targeted therapies were more likely to receive those treatments. Will we also find that patients who know about these important early detection methods are more likely to utilize them? Most likely.

For advocates to be helpful we do need to understand what the proteomics pipeline really contains and what the outcomes of the work have been in the past. Those outcomes certainly are more than just CA-125 and PSA but not much has been said in the patient literature. We need to understand what the barriers and promoters of the knowledge are along the way. Also, how the technologies and knowledge are being shared with other disciplines so that these methods can be integrated into clinical practice. We are also concerned about the rate of approvals as Henry mentioned even while discovery is higher.

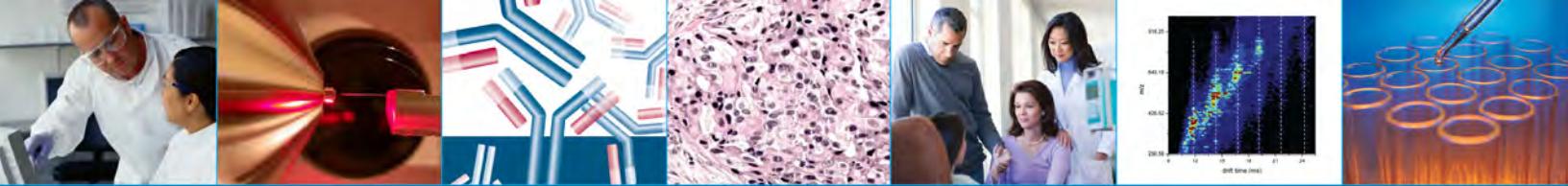
The CPTC is to be applauded for reaching out to advocates through this teleconference. We hope that there will be future opportunities to partner with advocates to truly fulfill the promise of proteomics and “team science.” We hope that advocates that represent the patient perspective will be considered an integral part of the “team” not by trying to be scientists but to partner together to participate in prioritizing discussions and problem solving, to review educational materials and to serve as a communications channel to disseminate information about the promise and realities of proteomics to our constituencies.

**Jeffrey Kaufman**  
Co-Founder and Executive Director  
Adenoid Cystic Carcinoma Research Foundation  
*eProtein Issue 2*

### **Lack of Metrics and Standards in Proteomic Discovery Technologies Hinders Innovative Research for Rare Diseases**

The personal experiences of Marnie Kaufman and her husband Jeff with adenoid cystic carcinoma (ACC) gave rise to the Adenoid Cystic Carcinoma Research Foundation (ACCRF). Mrs. Kaufman is a survivor of ACC—a rare, slow-growing cancer of the head and neck that typically originates in the salivary glands. Although the removal of her parotid gland and a course of radiation treatment sent her cancer into remission, the slow and persistent growth of ACC necessitates regular screenings for metastasis.

Dismayed by a lack of ongoing research in ACC, the Kaufmans decided to battle the disease on another front. In December 2005, they founded the ACCRF. Mr. Kaufman would eventually leave his role at a senior portfolio manager with Putnam Investors to take on the full-time job of Executive Director of ACCRF.



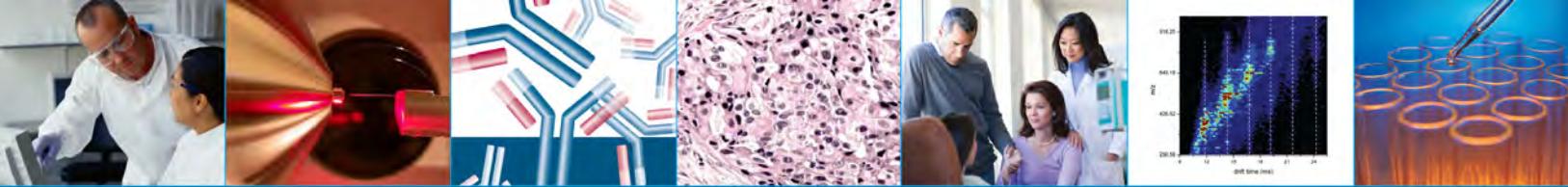
The mission of the ACCRF moves away from a traditional advocacy focus on patient support and education; the Foundation concentrates instead on coordinating ongoing research efforts to accelerate improvements in cancer treatment and the discovery of a cure. The approach of the ACCRF is simple: Advance discovery by identifying scientists and institutions working on the best research platforms and approach them with proposals for research partnerships. In doing so, the ACCRF seeks to establish a virtual network of ACC-focused researchers and clinicians to share resources and research findings.

During its short history, the ACCRF has developed a broad portfolio of research investments based on a wide array of platforms. Partnerships with the Sanger Institute, Göteborg University in Sweden, Harvard Medical School, and Johns Hopkins University have researchers developing a central biobank of ACC specimens, exploring gene mutations associated with ACC, and identifying gene targets through RNA interference techniques.

Despite the innovative and cutting-edge nature of the research funded by the ACCRF, its scientific advisory board has expressed hesitance in supporting proteomics-based research. “Due to concerns about the ability to replicate data collected through proteomics research, our organization has elected to focus our research investments in other areas,” says Mr. Kaufman.

The ACCRF is not alone in recognizing a lack of standards in proteomics research methods. The NCI attributes the vast discrepancy between the number of cancer protein biomarkers that have been described in scientific literature to date—over 1,200—and the surprisingly few that have transitioned into clinical applications to a lack of standardized procedures.

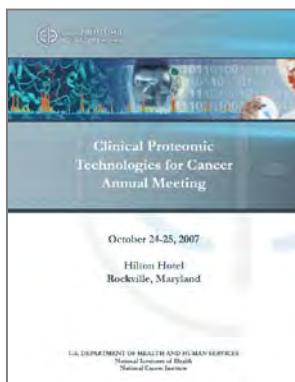
While ACCRF has not formally partnered with CPTC thus far, they are very encouraged by the work that the CPTC researchers are doing. “CPTC is providing a valuable service in addressing the challenge of validating the output of proteomics research,” Mr. Kaufman states. “I am confident that the work of CPTC will have an important impact on the future of proteomics research initiatives of the ACCRF.”



### 1.4.7.3 Annual Investigators' Meeting

#### ***First Annual Meeting***

CPTC held its first annual meeting in Bethesda, Md. on October 24-25, 2007, bringing together 170 participants from academia, government and industry. CPTC awardees presented updates on their progress in developing new tools, reagents, and technologies for protein and peptide measurement.

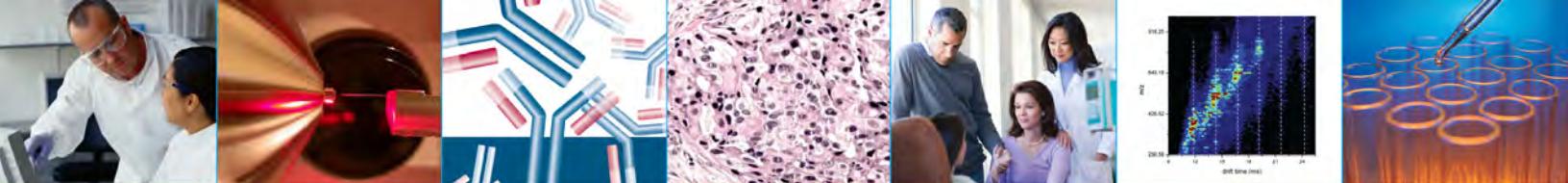


#### ***Second Annual Meeting***

CPTC held its second annual meeting in Cambridge, Mass. on October 28–29, 2008, bringing together more than 200 participants representing the full gamut of scientific fields that contribute to the initiative's mission to review the technological progress made over the previous year.

Giving a sense of the links between CPTC and other technology-focused initiatives supported by NCI, the first day of the meeting was held jointly with members of NCI's Innovative Molecular Analysis Technologies (IMAT) program. Several talks featured technologies and techniques developed by IMAT-supported investigators that have subsequently been applied to projects supported by CPTC, highlighting the importance of integrated technology development in cancer proteomics research in particular and in cancer research in general. The meeting also included talks and posters featuring research conducted through CPTC's three components: the CPTAC program, Advanced Proteomic Platforms and Computational Sciences, and the Proteomic Reagents and Resources Core.

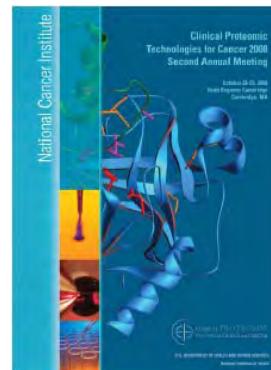
Both days featured keynote addresses by researchers speaking on their experiences in integrated research. David Altshuler, M.D., Ph.D., a founding member of the Eli M. and Edythe L. Broad Institute of MIT and Harvard and director of the institute's Program in Medical and Population Genetics, spoke of the lessons learned from conducting large-scale genomics research and how those lessons could apply to large-scale proteomics. In particular, he noted ways of avoiding pitfalls in validating variations, such as early development of robust, comprehensive, and scalable tools for determining



systematic associations—an issue that, he noted, CPTC is well on its way to addressing by focusing on technology development up front. Altshuler concluded his remarks by reminding attendees of the importance of grounding new discoveries in human biology before jumping to conclusions about their importance.

The second day's keynote, by Vamsi Mootha, M.D., of the Broad Institute and Massachusetts General Hospital, focused on integrative genomic, proteomic, and metabolomic research on mitochondrial diseases. The mitochondrial proteome has not yet been fully defined, but it may contain between 1,200 and 1,500 proteins, only 13 of which have been associated with genes found in the mitochondrial genome; the rest are encoded by nuclear genes. Mootha's talk outlined his work to develop a mitochondrial protein catalog, called MitoCarta, which currently contains 1,098 mitochondrial proteins. With this information in hand, he has started probing the ancestry of numerous mitochondrial proteins and applying that knowledge clinically to explore rare familial diseases caused by breakdowns in mitochondrial respiration oxidative phosphorylation.

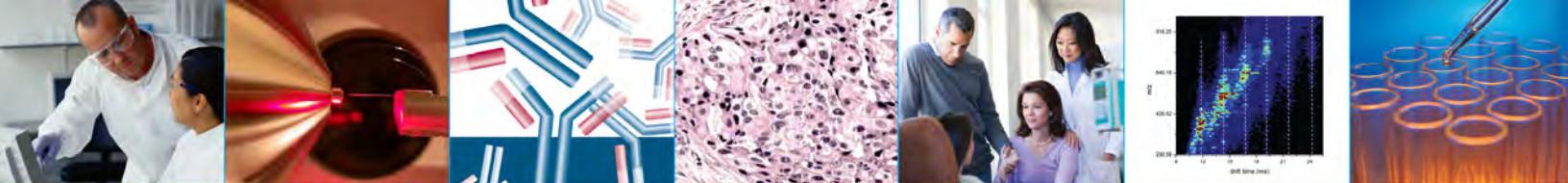
In his closing remarks, CPTC Director Henry Rodriguez, Ph.D., M.B.A., noted that the initiative had produced some very good outputs since its launch two years ago. Rodriguez also mentioned that while there had been a learning curve associated with the initiative, they had shown that team-based science can be very successful, and that the steps that had been undertaken thus far had laid the groundwork for CPTC's future success.



The annual investigator's meeting agendas can be found in Appendix 4.8.

#### 1.4.7.4 NCI-NIST Next-Generation Methods of Peptide Identification Workshop

In November 2007, CPTC Program Staff partnered with NIST to host a workshop on Next-Generation Methods of Peptide Identification. The goal was to bring together thought leaders and stakeholders in the proteomics community to assess the status of peptide identification – a fundamental task for all mass-spectrometry-based proteomics. Over 60 experts from around the country attended and provided lively discussion around the evaluation of several methods. Techniques considered ranged from traditional



database sequence searching and sequence tag searching to avant garde methods of spectral libraries and theoretical fragmentation prediction.

#### 1.4.7.5 International Proteomics Data Release and Sharing Policy Summit: Amsterdam Principles

Data sharing is standard practice among members of the genomics community, based on principles developed at a 1996 gathering in Bermuda and ultimately endorsed by all major parties in the Human Genome Project. The widespread sharing of prepublication sequence data greatly accelerated the pace of genomic discovery. However, similar policies do not exist for proteomics research, a state of affairs currently seen as a significant obstacle to progress in the field.

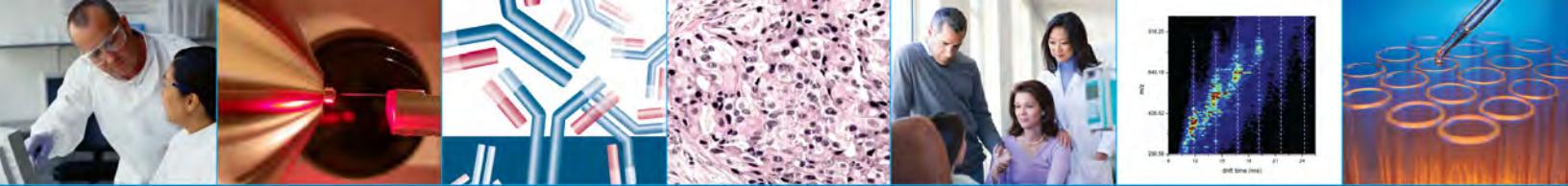
In August 2008, the NCI sponsored a summit in Amsterdam for members of the international proteomics community, including representatives from funding agencies, journals, and academic research centers. Their task: to begin defining policies and practices that would govern and facilitate the release of proteomic data into the public domain along the Bermuda model. The summit's participants addressed the following questions:

**When should data be released?** Participants agreed that the timing of data release should be governed by the type of project. Investigators taking part in community resource projects should be required to release data once they are produced. Investigators working on individual projects, on the other hand, should release data upon publication in a peer-reviewed journal.

**What types of data should be released, and what kinds of metrics should be used to define data quality?** Participants agreed that high quality, well-annotated raw data (for MS and protein/affinity array data) would be the most reliable interchange format for data repositories. Metadata, information on data quality, and identification quality control will all be critical as well. Accessing these data would require development of the proper infrastructure (i.e., community supported standardized formats, controlled vocabularies and ontologies, minimal reporting requirements, and publicly available online repositories). Central repositories should develop their own thresholds for data quality metrics, in a coordinated manner with users and one another, to ensure interoperability.

To fuel progress in proteomics research, data sharing cannot be voluntary; rather, it is up to scientists, journals, and funding agencies to take the necessary steps to ensure that all parties adhere to the standards for data release, ideally within a frame work of tripartite responsibility akin to that created for genomics research. Central repositories, for their part, should clearly define minimum submission requirements, encourage rich annotation, and develop seamless submission procedures.

A white paper based on the discussions of the summit is forthcoming.



#### 1.4.7.6 Alignment with FDA Critical Path



The Food and Drug Administration (FDA) launched the Critical Path Research Initiative to identify, develop, and apply state-of-the-art genomics and proteomics technologies to medical product development to improve the accuracy of the tests used to predict the safety and efficacy of investigational medical products.

In alignment with the Critical Path Initiative, a memorandum of understanding (MOU) between the FDA and the NCI became effective on April 5, 2007. The purpose of this MOU was to establish a formal collaboration regarding proteomics science and technology to accelerate proteomics technology development and application in clinical settings. FDA and NCI are collaborating in areas involving proteomics such as: Sample collection, preparation, storage and processing; bioinformatics and data analysis; discovery and validation of biomarkers; and surrogate biomarkers of cancer development and drug response, including standardization among technology platforms and assay standards development.

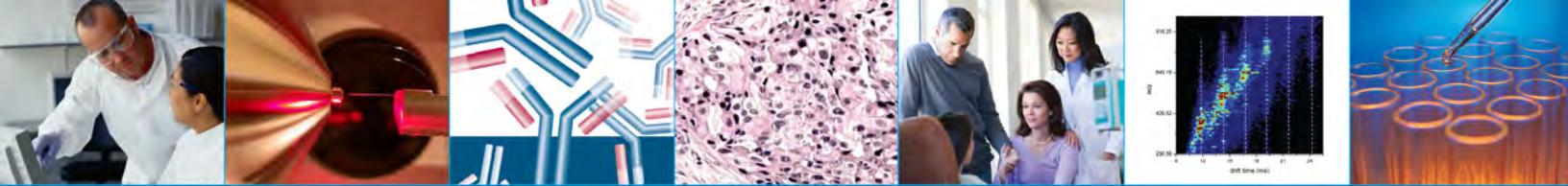
The MOU between FDA and NCI can be found in Appendix 4.9.

#### 1.4.7.7 NCI-FDA Interagency Oncology Task Force Molecular Diagnostics Workshop



The NCI-FDA Interagency Oncology Task Force (IOTF) Molecular Diagnostics Sub-Committee was established in May 2007. The goals of the Sub-Committee are to:

- Identify efforts that have the most potential (guidance documents, educational material, Critical Path projects)
- Develop analytical validation guidelines to define and illustrate the principles that researchers developing protein-based assays should address to assure measurement accuracy and reliability.



## NCI-FDA IOTF Molecular Diagnostics Sub-Committee Members

*Co-Chairs:* Henry Rodriguez (NCI), Elizabeth Mansfield (FDA)

*Members:* Estelle Russek-Cohen (FDA), Gary Kelloff (NCI), James Jacobson (NCI), Larry Kessler (FDA), Mark Raffeld (NCI), Mitch Gail (NCI), Ruth Pfeiffer (NCI), Steve Gutman (FDA), Zivana Tezak (FDA)

An IOTF Molecular Diagnostics Workshop was held on October 30, 2008 in Cambridge, MA. The purpose of this workshop was to discuss requirements for analytical qualification of proteomic technologies (MS and affinity arrays) in the context of its intended use. This workshop focused on technology-specific analytical validation processes to be addressed prior to use in clinical settings. Focused discussion areas included sources of assay variability, selection of standards/quality control materials, application of internal and external control strategies, and appropriate metrics for quality control, instrument and assay performance analyses.

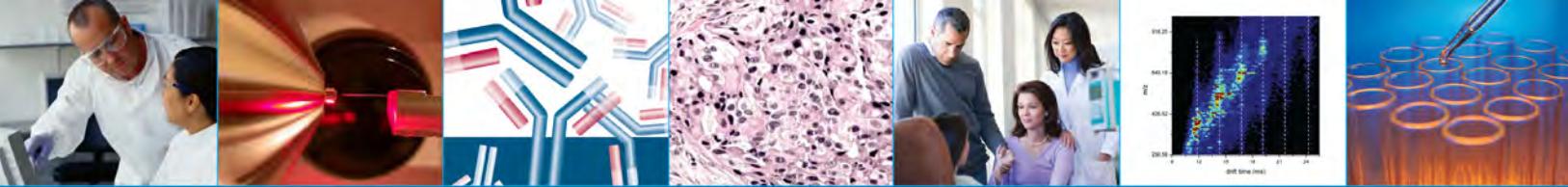
The workshop used a case study approach that discussed issues related to:

- Qualifying a proteomic technology;
- Specimen and population issues;
- Statistical issues; and
- Understanding the regulatory pathway to commercialization.

The agenda is available in Appendix 4.10.



A summary document is in development for publication, which will discuss analytical validation issues that specific proteomic technologies should address when seeking FDA approval. In addition, a mock 510(k) pre-application filing is being considered. This mock filing would help orient the FDA to MS in novel diagnostics and serve as a springboard for guidance to the MS proteomics community.



#### **1.4.7.8 National Heart, Lung and Blood Institute**

In their search for valid biomarkers for acute coronary syndromes, scientists from the National Heart, Lung and Blood Institute (NHLBI) are adopting verification technology introduced by CPTAC scientists that bridges biomarker discovery and clinical validation. For greater detail on the adoption of this technology by NHLBI, please see Appendix 4.11.

#### **1.4.6.9 American Association for Clinical Chemistry**

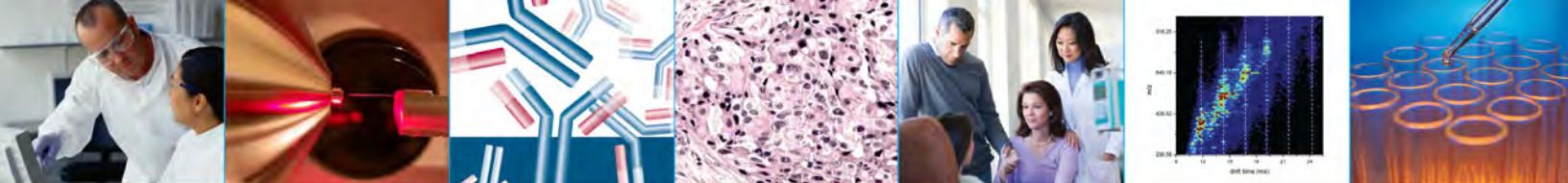


The AACC (American Association for Clinical Chemistry) is an international society comprised of medical professionals with an interest in clinical chemistry, clinical laboratory science, and laboratory medicine. Founded in 1948, the society has over

9,000 members and is headquartered in Washington, DC. AACC recently established a proteomics division in order to educate laboratory professionals about the diagnostic applications of proteomics and to promote the use of proteomics-based tools in medical care. The AACC Proteomics Division thus provides a forum for sharing knowledge, ideas, experience, and strategies for diagnostic applications of proteomic methods.

CPTC and the AACC Proteomics Division agree that it is in the best interest of both organizations to develop a partnership that enhances and enables the proteomics community to unite and battle cancer. As a result, the CPTC have joined forces on educating the clinical chemistry community in the area of proteomic standards and technology advances. Examples of such exchanges include:

- Development of editorials and special issues in the journal of *Clinical Chemistry* on clinical proteomic standards and methodologies.
  - CPTC participation at AACC Proteomics Division annual meetings and AACC annual meetings
  - AACC participation at CPTC governing body meetings and CPTC annual meetings



#### 1.4.7.10 Korean Functional Proteomics Center



In July of 2002, the Korean Ministry of Science and Technology established the Functional Proteomics Center (FPC) - one of its 21<sup>st</sup> Century Frontier R&D Initiatives. The Center's goal is to establish an infrastructure of proteomics core technology in Korea with a mission of identifying novel biomarkers and therapeutic target proteins for human diseases. The research teams in the program are to focus on technology development, as well as on proteomic analysis of samples from disease models and appropriate patients, which will be followed by studies on protein networks and disease mechanisms.

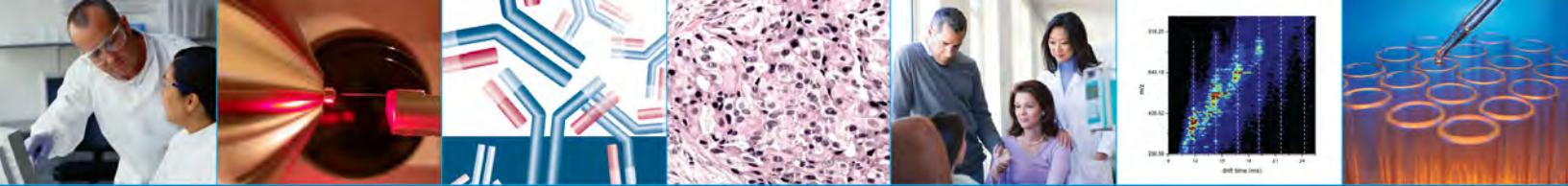
CPTC and FPC agree that it is in the best interest of both organizations to develop a strategic partnership that enhances and enables the international proteomics' community to unite and battle cancer. Over the past two years, both initiatives have worked closely in the exchange of scientific knowledge and technology/standards development.

Examples of such exchanges include:

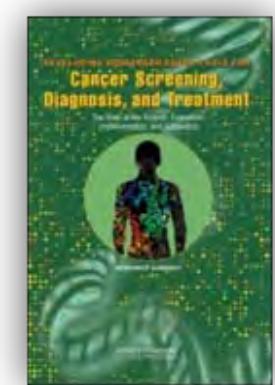
- Adoption of NCI's CPAS software by the Korean Functional Proteomics Center  
CPAS is a set of web-based bioinformatics and collaboration tools to help scientists store, analyze, and share data from high-throughput experiments and clinical trials. CPAS has recently been approved as caBIG Gold compliant through efforts by CPTC.
- Introduction of MRM technology  
MRM technology is being introduced by NCI's CPTC to the clinical proteomics community. CPTC is empowering laboratories with this technology by the development of robust methodologies and standards. CPTC is working with the FPC in having MRM technology, as part of a Verification Phase, introduced into their clinical research centers.

#### 1.4.7.11 Biotechnology and Pharmaceutical Industry

A number of articles have been written for trade publications that target the biotechnology and pharmaceutical industries. These include, but are not limited to, Scientific American worldVIEW, Genetic Engineering News, Drug Discovery News. Article reprints can be found in Appendix 4.12.



#### 1.4.7.12 Institute of Medicine of the National Academies



Institute of Medicine (IOM) Report: “*Current proteomic technology approaches are insufficient to reliably and reproducibly discover, identify, and quantify peptides and proteins of clinical significance for cancer*” from complex patient samples.

A long-standing goal in cancer biology has been to develop tests that can detect cancer early, accurately predict prognosis, and facilitate selection among therapies. The use of biomarker tests could also greatly facilitate and accelerate the development of targeted cancer therapies by helping companies choose the most promising drug candidates and by identifying patients that are most likely to benefit from a given therapy.

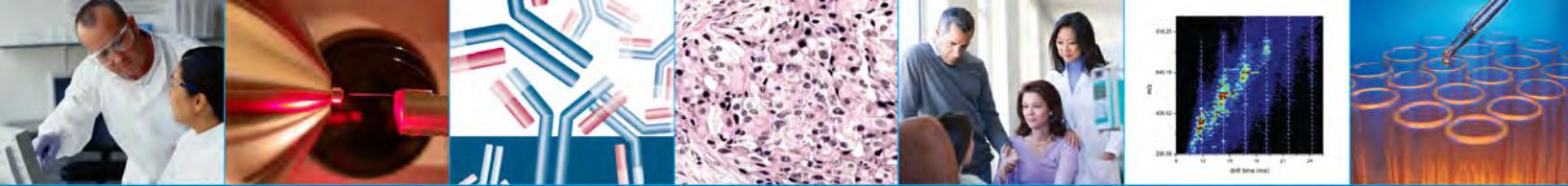
Recent technological advances, especially in the fields of genomics and proteomics, have made it easier to identify many biomarkers at once in high-throughput screens. Such advances have fueled interest in this research area, although validation of markers - that is, determination of clinical relevance and applicability - is quite challenging, and many questions have been raised regarding how new tests will be developed, evaluated, and integrated into clinical practice.

Nonetheless, widespread adoption of effective new biomarker tests for cancer detection and diagnosis as well as therapy selection and monitoring may lead to a paradigm shift in the way that medicine is practiced, with potential economic consequences.

In 2006, an ad hoc committee, via a workshop, examined questions regarding:

- the potential to improve cancer screening, diagnosis, and therapy through the use of emerging biomarker technologies;
- current limitations of genomics and proteomics technologies for cancer detection, diagnosis, and drug development, and steps that could be taken to improve them;
- the logistics and cost of coordinating the development of biomarkers and targeted therapies;
- regulatory oversight of biomarker development and use;
- the adoption of biomarker-based tests and therapeutics into clinical practice; and
- some of the potential economic implications of adopting these emerging technologies.

CPTC is addressing issues pertaining to proteomic technologies.

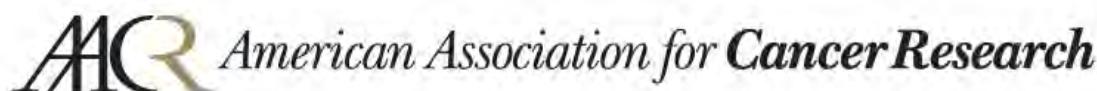


### *About Institute of Medicine*

The nation turns to the IOM of the National Academies for science-based advice on matters of biomedical science, medicine, and health. A nonprofit organization specifically created for this purpose as well as an honorific membership organization, the IOM was chartered in 1970 as a component of the National Academy of Sciences.

The Institute provides a vital service by working outside the framework of government to ensure scientifically informed analysis and independent guidance. The IOM's mission is to serve as adviser to the nation to improve health. The Institute provides unbiased, evidence-based, and authoritative information and advice concerning health and science policy to policy-makers, professionals, leaders in every sector of society, and the public at large.

### **1.4.7.13 American Association of Cancer Research (AACR)**



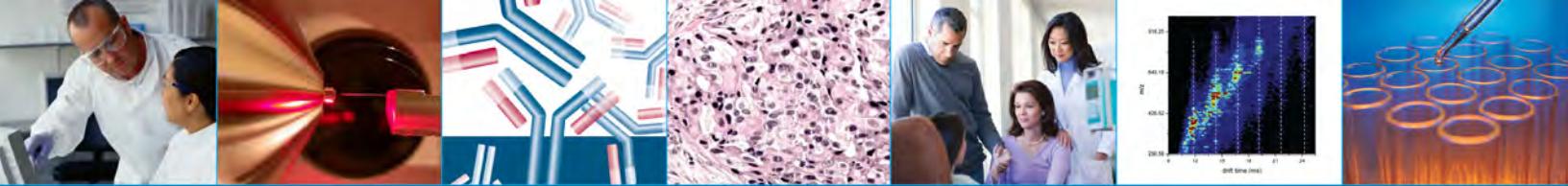
The CPTC program, through its CPTAC Center Network, has emphasized the development and implementation of standardized proteomic platforms for discovery and verification studies, the use of performance standards and performance metrics and the evaluation of technology platforms through SOP-driven studies.

Because the mission of the AACR is to prevent and cure cancer through research, education, communication, and collaboration, the CPTC hosts special educational/methods workshops at their annual meetings. These workshops discuss the challenges inherent in developing proteomic biomarkers for cancer; the approaches of the CPTACs to implementing, standardizing and validating analytical proteomics platforms; the application of proteomics technologies to translate biomarker candidates discovered in mouse models to human cancers; and, the discovery and verification of biomarkers entirely in a human cancer context. The first workshop was held in 2009 and was standing-room only audience. CPTC is working with the members of AACR to ensure future events.

### **1.4.7.14 Human Proteome Organization (HUPO)**



The Human Proteome Organization (HUPO) is an international scientific organization representing and promoting proteomics through international cooperation and collaborations by fostering the development of new technologies, techniques and



training. Since 2006, CPTC has been invited every year to host special methods and standards workshops. The purpose of these sessions is to discuss current CPTC findings to their members.

## 1.5 Leveraged Funding

### Broad Institute of Harvard and MIT

**University of Washington:** Established a collaboration with Dr. Andy Hoofnagle (Dept of Laboratory Medicine at the Univ. of Washington, Seattle), developer of a clinical SISCAPA assay against thyroglobulin (used clinically as a recurrence marker for thyroid cancer). Using our reagent pipeline, the Broad team is preparing improved anti-peptide antibodies to extend the sensitivity of his SISCAPA assay to match the sensitivity of currently-available immunoassays (which suffer from serious interference inaccuracies). This effort will result in a SISCAPA assay that can be introduced into clinical practice at UW, allowing us to gain experience in the routine use of this format in a practicing clinical lab.

**NHLBI:** With Drs. Robert Gerszten, MGH and Marc Sabatine, HMS, the Broad team employed SISCAPA technology developed under the auspices of the NCI CPTC program to construct the first multiplexed SISCAPA assay for two clinically relevant proteins (cTnI and IL-33) in cardiovascular disease as part of an NHLBI-funded cardiovascular biomarker project. Importantly, this is the first assay for IL-33 with potential clinical utility.

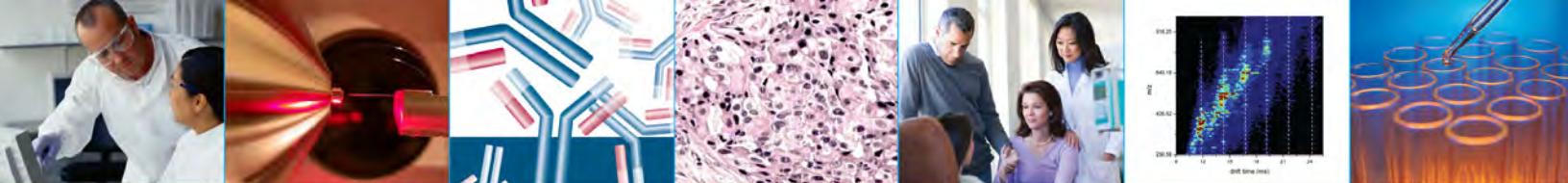
**University of Victoria:** With Dr. Christoph Borchers, University of Victoria, the team has helped develop a collection of 45 stable isotope labeled tryptic peptides representing the 45 most abundant proteins in human plasma. These reagents will be used in CPTAC-wide studies to assess and optimize conditions for proteolytic digestion, a key step in all proteomics experiments that remains poorly understood.

**Women's Cancer Research Fund:** Collaborated with the Women's Cancer Research Fund to use breast cancer proteomics data for assay development in CPTAC-wide inter-laboratory studies of MRM and SISCAPA-MRM performance and reproducibility.

Four externally-funded biomarker assay projects are making use of the SISCAPA reagent pipeline developed by NLA and TWP for the Carr CPTAC program. The projects are managed by Anderson Forschung Group LLC (AFG) and the Plasma Proteome Institute (PPI). They include:

**The Canary Foundation:** funded development of SISCAPA reagents for two biomarker targets to inaugurate the multiplex-immunization antibody pipeline developed with Epitomics, Inc.(managed by PPI).

**Pfizer Inc:** has funded a pilot collaboration with AFG to develop SISCAPA assays to 10 candidate markers selected by Pfizer. The results of the completed project were successful and demonstrate that SISCAPA assays can be multiplexed. Joint publications describing this work are being prepared (Managed by AFG).



**The Translational Genomics Institute:** (Phoenix, AZ) has funded a pilot collaboration with AFG to develop SISCAPA assays to 3 candidate autism markers selected by TGEN (Managed by AFG).

**biOasis Technologies Inc.:** has engaged the AFG to develop a SISCAPA assay to measure the levels of p97 (melanotransferrin; a potential marker for Alzheimers disease) in human blood samples (Managed by AFG).

### University of California at San Francisco

**The Canary Foundation:** As a direct result of our participation in the CPTAC program, the team received a \$600,000 grant from the Canary Foundation to purchase an Applied Biosystems 4000 mass spectrometer.

### Purdue University

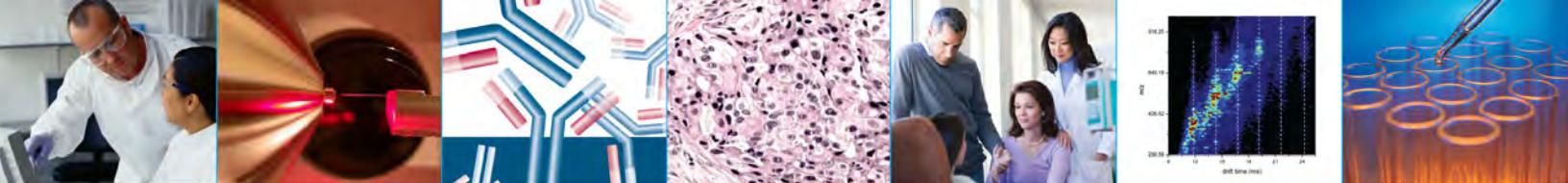
#### Establishment of a New Research Institute: The Center For Analytical Instrumentation Development (CIAD)

Co-directors: Graham Cooks and Fred Regnier



The NCI sponsored Cancer Center at Purdue is among the oldest in the U.S. Moreover, Purdue has a strong Oncological Sciences program initiated in 2004 with funding from Eli Lilly. These Centers were joined in 2006 by the NCI funded Clinical Proteomics Technology Assessment for Cancer (CPTAC) Center, and now form the nucleus of a strong program for the evaluation of cancer diagnosis and treatment. CAID brings together scientists from Purdue, Indiana University (IU), Indiana University School of Medicine (IUSM), and the University of Illinois (UIUC) for the development next generation clinical instrumentation. Commercialization of instrumentation will be aggressively pursued to assure broad availability of new technology.

Instrumentation developed in this program is providing economic opportunity for the region through 1) vigorous commercialization efforts by the Purdue Office of Technology



Commercialization (OTC), 2) prototyping (early manufacturing) of alpha- and beta-version instrument platforms in Bindely, and 3) introduction of prototype instruments into the life science community before full commercialization. As an example, BioChip based immunological assays are an area where miniaturization of Purdue technology is doing well. More than a million canine heart worm assays were carried out across the U.S. on these chips in 2008. Both of these technologies could be of substantial efficacy in human diagnostics.

### New NIH Funding for Validation of Protein-Based ‘Wellness’ Test

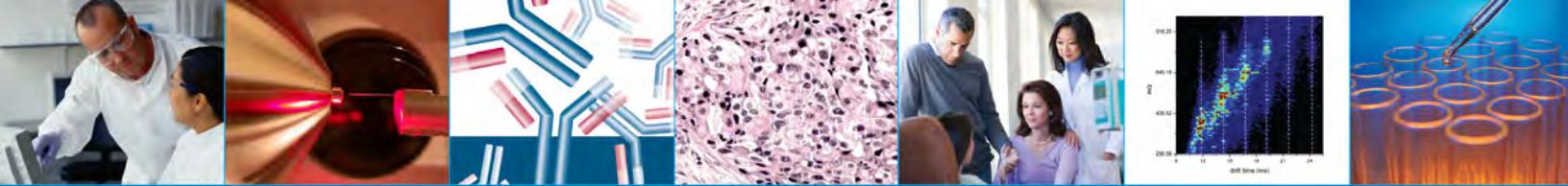
Involvement of Purdue University as an integral part of CPTAC helped Predictive Physiology & Medicine (PPM), a spin-out company from Indiana University, to secure a two-year, \$2.34 million grant from the National Institutes of Health to validate a proteomics- and metabolomics-based testing platform it is developing. The PPM platform is designed to test protein and metabolite biomarkers and other elements in patient blood samples to yield an overall health evaluation. That evaluation is then packaged into a wellness test, made available to patients through their doctors.

### Cancer Care Engineering

Cancer Care Engineering is a systems engineering approach to colon cancer research and healthcare delivery with \$5 million in funding over two years. The award is to the Oncological Sciences Center, which is partnered with Indiana University School of Medicine. The project is applying a systems (biology) approach on the research side with major emphasis on integration of disparate data types obtained from genomic, proteomic and metabolomic analyses of prospectively collected colon cancer patient samples.

The website [ccehub.org](http://ccehub.org) is the core IT resource for the project that utilizes the Purdue HubZero technology to create a user environment for information sharing, resources including modeling and simulation, data analysis capability (our LCMS analysis pipeline) as well as a data repository.

The Bindley Bioscience Center, in Discovery Park (Purdue University), is creating and making available the proteomics and metabolomics data from these samples (including both profiling and glycoprotein proteomic analyses). This Cancer Care Engineering project was enabled by Purdue’s participation in CPTAC. (The modeling and data visualization approaches developed and used in the CCE project will also be applied to the CPTAC breast and prostate data.)



## 1.6 Public-Private Partnerships

### 1.6.1 Small Business Innovation Research

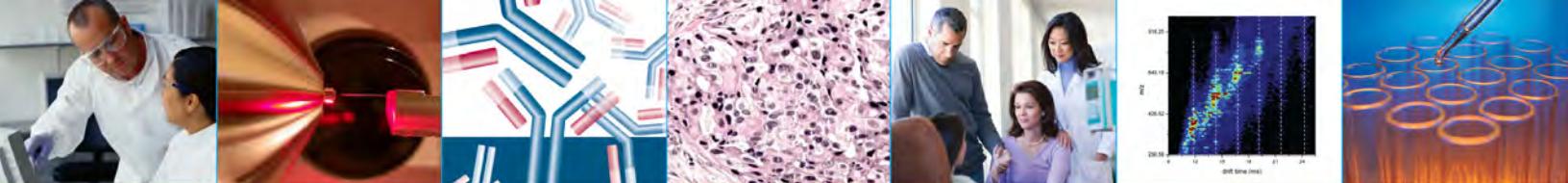
Ready access to high-quality, standardized reagents is of great importance if the proteomics community is to catalyze biomarker discovery for reducing the burden of cancer.



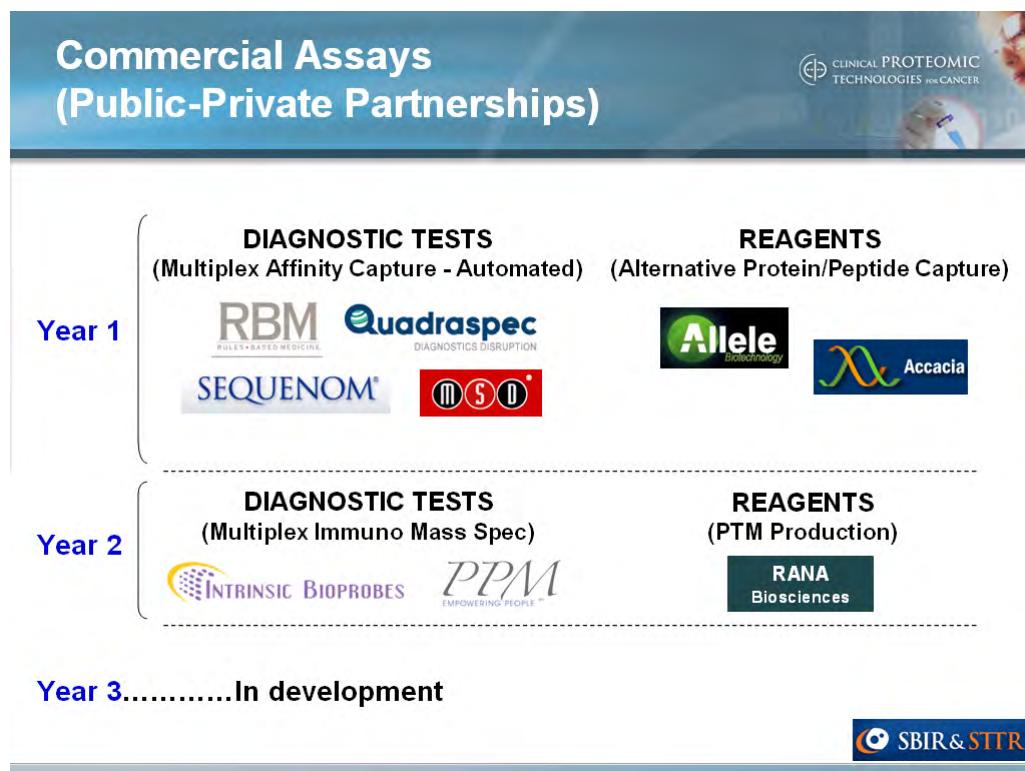
Numerous small businesses design and develop proteomic technologies for the accurate and powerful measurement of proteins and other biomolecules related to disease. Without well-characterized reagents, however, it is impossible to translate such platforms into products and services that could be used effectively by the cancer community.

One of CPTC's three component programs, the Proteomic Reagents and Resources Core, is tasked with providing the cancer community with the tools necessary to overcome technological and methodological barriers to developing and providing such reagents. To maximize the Core's capabilities and impact, CPTC is partnering with the biotechnology industry via the NCI's Small Business Innovation Research (SBIR) Program, a contract mechanism that supports early stage research and development by small businesses. Through the SBIR program, CPTC aims to integrate its efforts with those of the biotechnology industry by encouraging and enabling companies developing proteomic technologies and platforms to adopt standardized, well-characterized reagents—including high quality proteins and validated capture reagents (e.g., antibodies)—in the commercialization of new tools and kits for the cancer community.

CPTC has awarded nine contracts based on SBIR requests for proposals released in 2007 and 2008 (Figure 8). For fiscal year 2009, CPTC sought proposals on "Novel Antibody Epitope Mapping Technologies," "Development of Novel Protein Expression Technologies for Glycosylated Cancer Related Proteins," and "Peptide Aptamers: New Tools to Capture and Study Protein Interactions in Lieu of Immunological Reagents." Contract awards for these topics are anticipated to be announced in the summer of 2009.



**Figure 8. 2007 and 2008 SBIR Contract Recipients**

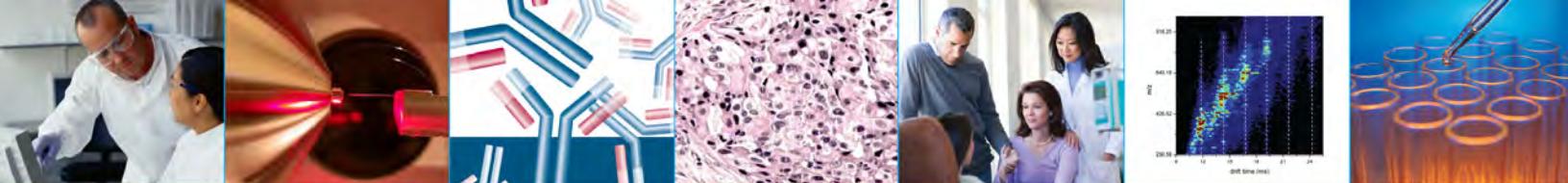


### 1.6.2 Innovative Molecular Analysis Technologies

The NCI-sponsored Innovative Molecular Analysis Technologies (IMAT) program and the CPTC program possess the mutual aims of developing, integrating, and validating novel and emerging technologies in support of cancer research, diagnosis, and treatment.



The IMAT program is aimed at the development, technical maturation, and dissemination of highly innovative technological platforms having a high degree of technical risk but also the potential to significantly impact and transform the quality and depth of cancer research. Technologies developed through IMAT are suitable for the molecular analysis of cancers in multiple environments, including basic, clinical, and epidemiological research settings.



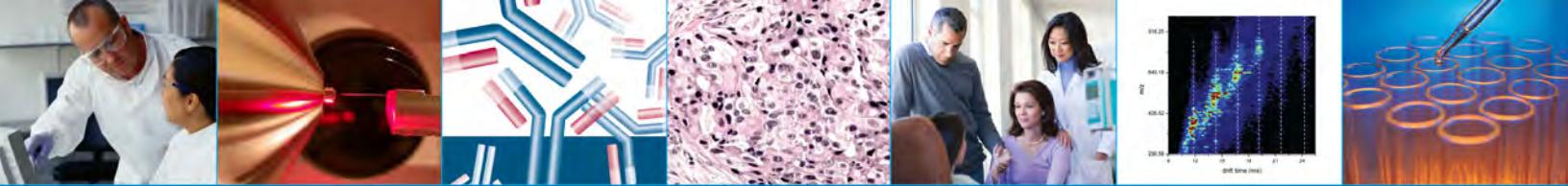
A significant, if not persistent, obstacle to the development and use of innovative advanced molecular platforms has been the lack of access to well-characterized and validated affinity reagents and supporting resources by which to adequately test the platform. In order to empower the sustained maturation of innovative technologies, access to well-characterized, well-validated reagents capable of accelerating biomarker discovery and validation, therapeutics monitoring, and cancer diagnostics development, is critical. Such reagents have recently been launched and made available to the general community by the CPTC program. The CPTC Proteomic Reagents and Resources Core includes highly characterized monoclonal antibodies to human proteins associated with cancer, labeling reagents, protein and/or peptide mixtures, and other reagents needed for effective proteomic analytical platform characterization.

In 2009, qualified innovative analytical technologies that have passed through IMAT's initial phases of technical development and that possess a demonstrated record of feasibility as established through the IMAT program, but that also require the use of monoclonal antibodies or other such resources, will be solicited as part of a joint IMAT-CPTC effort to further validate technological parameters such as utility, capability, breadth, and relevance by leveraging and incorporating well-established and well-characterized monoclonal antibodies developed through the CPTC Reagents & Resources Core. This builds upon the sustained interaction and partnership between both programs.

In October 2008, the IMAT and CPTC programs held their first-ever back-to-back Principal Investigators meeting in Cambridge, Massachusetts. The meeting succeeded in fostering cross-collaborations among different research groups and between individual investigators from both programs. Among the highlights of the meeting was a joint scientific session in which investigators from both initiatives explored and highlighted their mutual areas of need and potential synergies. Specific examples of cross-over technologies, or platforms that were developed through IMAT and subsequently implemented through CPTC, were highlighted. A similar joint meeting is currently envisioned for October, 2009. This latter meeting will continue to build upon the groundwork established at the 2008 meeting and will include an expanded joint scientific and poster session as well as networking opportunities.

### 1.6.3 University of Iowa Developmental Studies Hybridoma Bank

The Developmental Studies Hybridoma Bank at the University of Iowa (DSHB) was created by NIH as a national resource, and has been given the distribution rights for monoclonal antibodies being generated by NCI biomarker-related initiatives. DSHB at the University of Iowa collects, stores, and distributes fully-characterized antibodies and hybridomas generated by CPTC. Researchers are charged a small fee.



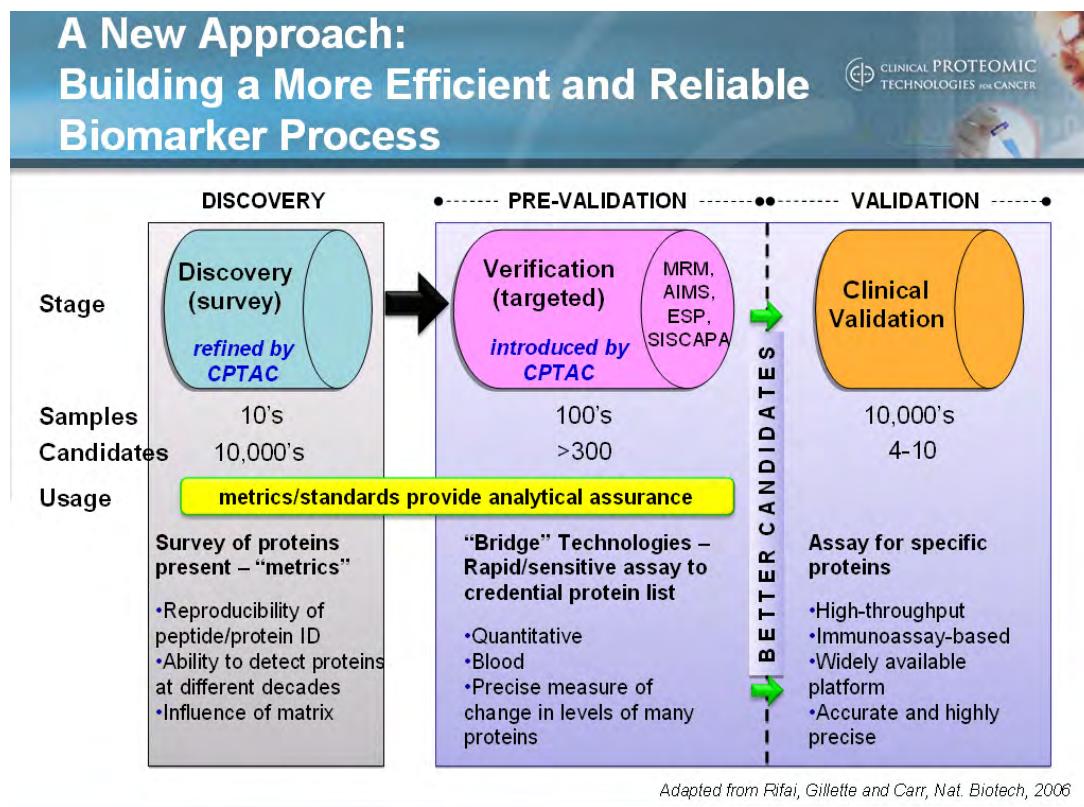
## 2. Program Performance

### 2.1 Introduction

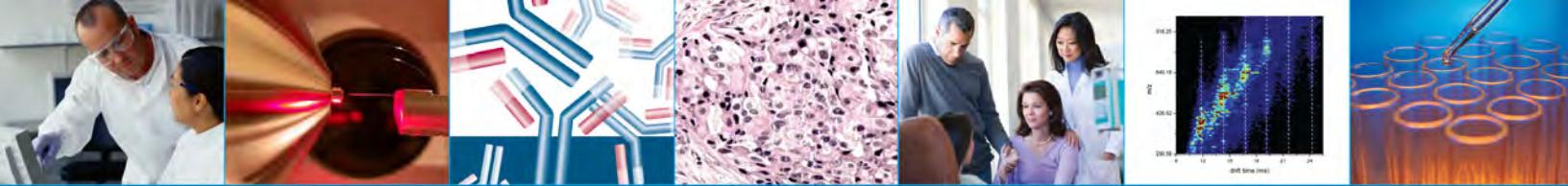
CPTAC represents one of the most in-depth multidisciplinary networks to optimize existing proteomic analysis platforms to reliably identify, quantify, and compare peptides/proteins in complex biological mixtures. CPTAC has conducted a number of “round robin” studies designed to accomplish two over-arching goals (Figure 9):

1. Refine the biomarker discovery pipeline
2. Introduce a pre-validation step, verification, into the pipeline using “bridge” technologies

**Figure 9. CPTAC is Building a Better Biomarker Development Pipeline**



The following section highlights the results of these studies. A complete list of publications and patents can be found in Appendix 4.13 and 4.14, respectively.



## 2.2 Mass Spectrometry Platform Benchmarking “Round Robin Evaluations”

### 2.2.1 Multiple Reaction Monitoring Performance Assessment

Biomarker discovery studies typically result in a lengthy list of tens to hundreds of potential candidates, most of which are false positives. Attempting to identify the “gems” in lengthy lists—those biomarkers that are sufficiently sensitive and specific for cancer—has proven to be particularly challenging and remains the rate-limiting step for clinical translation. Candidate-based biomarker verification is increasingly being viewed as a critical step in the biomarker discovery pipeline, bridging unbiased discovery to pre-clinical validation.

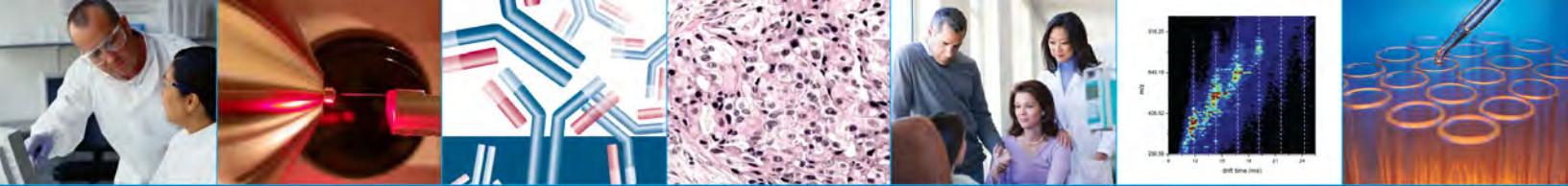
In the past, verification of novel biomarkers has relied on the use of high-throughput immunoassays; however, the development of these assays relies on the availability of suitable, well-characterized antibodies. Furthermore, such reagents for novel biomarkers do not exist and the time, expense, and technical limitations required to generate them provide a strong incentive to develop alternative approaches.

Targeted MRM assays have been successfully used to quantify small molecules in pharmaceutical research as well as in clinical laboratories (e.g. newborn screening). Recently, several individual laboratories have successfully used MRM coupled with stable isotope dilution mass spectrometry (SID-MS) to directly quantify proteins in human plasma. SID-MRM-MS simultaneously targets and measures a moderate amount of candidate proteins through detection of “signature” peptides that are diagnostic for each protein. Using this method, it becomes feasible to rapidly screen large sample banks in pre-clinical studies to obtain statistically valid data for verification purposes.

However, the reproducibility and transferability of protein-based MRM assays across different instrument platforms and laboratories has yet to be demonstrated. CPTAC has addressed this issue through large, multi-site, consortium-wide studies designed to evaluate intra- and inter-laboratory analytical performance of SID-MRM-MS assays. The results of these studies include the following highlights:

- A protocol has been successfully designed and implemented to measure absolute amounts of seven proteins spiked into human plasma, providing a foundation for the proteomics community for the development of similar protein quantitation methods.
- The reproducibility of protein quantitation across multiple sites with differing LC-MS platforms has been demonstrated.
- The greatest source of variation lies with sample processing.

Beyond biomarker candidate verification, SID-MRM-MS assays are also being considered as potential replacements for several clinical immunoassays due to the greater molecular (structural) specificity of MRM measurements as compared to antibody:antigen binding and the relative ease of multiplexing. The simplicity of producing and characterizing peptide-based reference materials for SID-MRM-MS will

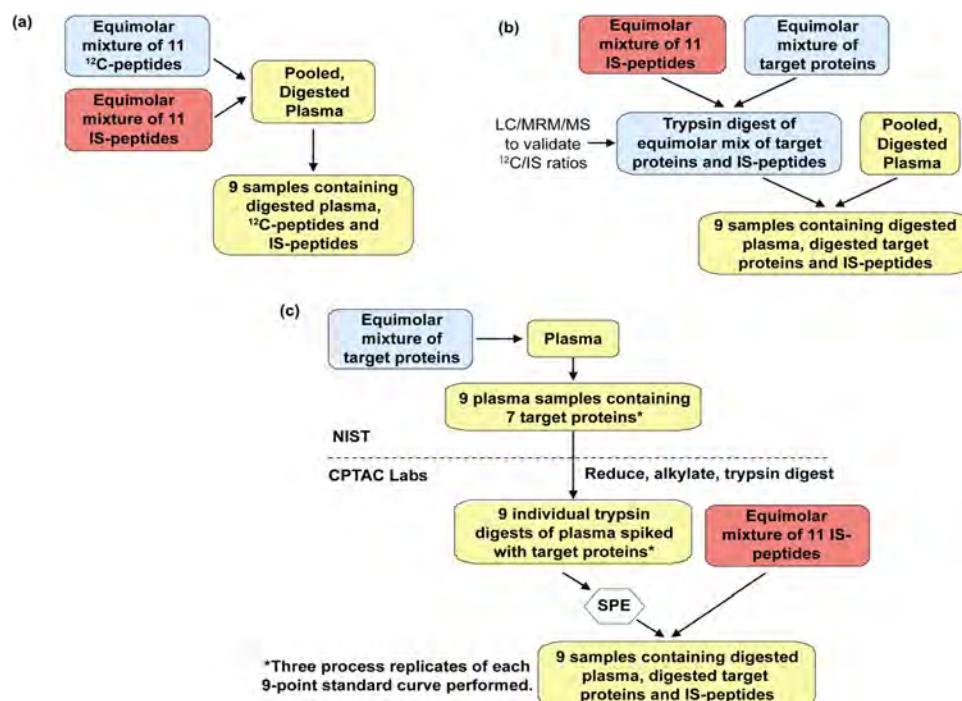


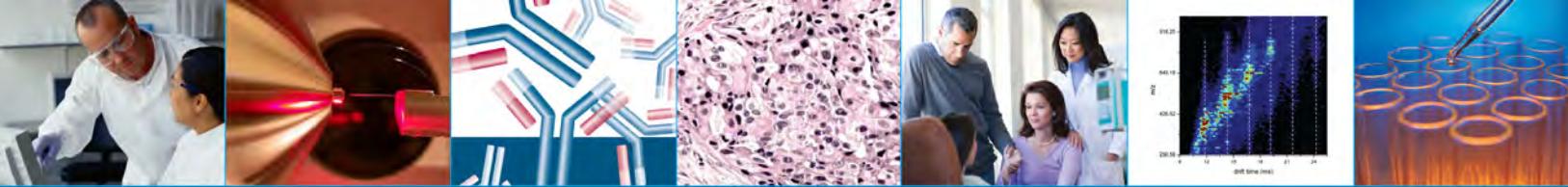
overcome the known issues of ELISA assay standardization that lead to varying results across multiple clinical laboratories and will provide a critical component for a systematic biomarker pipeline.

**A Multi-site Assessment of Precision and Reproducibility of Multiple Reaction Monitoring-based Measurements: Toward Quantitative Protein Biomarker Verification in Human Plasma. Accepted for publication in Nature Biotechnology, July 2009.**

**Figure 10. Sample preparation workflow for Studies I, II, and III**

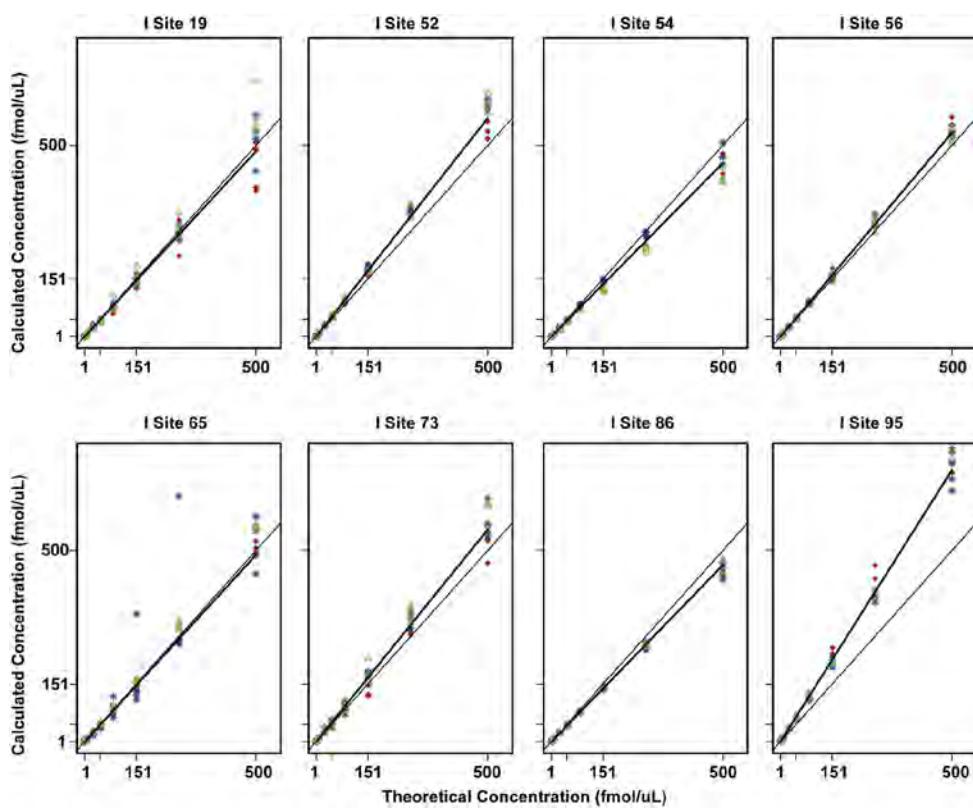
Workflow (a) represents Study I:  $^{12}\text{C}$  and  $^{13}\text{C}/^{15}\text{N}$  peptides were spiked into pooled, digested plasma to generate a 9-point standard curve. Workflow (b) represents Study II: an equimolar mixture of the seven target proteins was digested separately and spiked into pooled, digested plasma and diluted with a mixture of IS peptides (prepared in pooled, digested plasma) to generate a 9-point standard curve. Study I and Study II samples were prepared centrally at NIST. Workflow (c) represents Study III: an equimolar mixture of the target proteins was spiked into neat plasma and diluted with plasma to generate a 9-point standard curve. Three aliquots of these samples (prepared at NIST) were then shipped to the eight participating CPTAC sites where reduction, alkylation, digestion and desalting were carried out prior to SID-MRM-MS analysis.

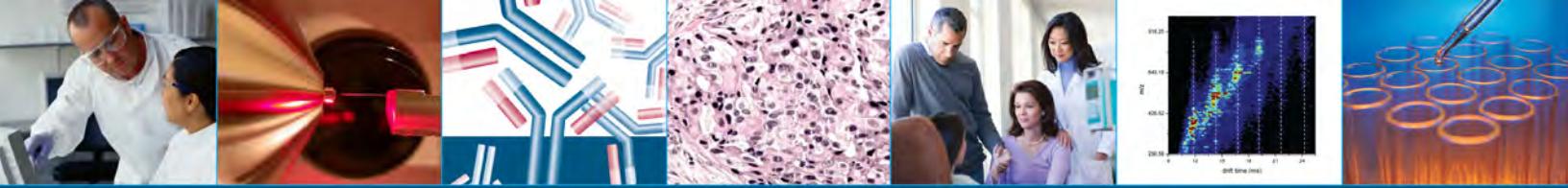




**Figure 11. Inter-laboratory reproducibility of linear calibration curve slopes for Study I**

The eight plots display the concentration curves for APR-AGL in Study I across all labs. Each of the eight sites were assigned random numerical codes (19, 52, 54, 56, 65, 73, 86, 95) for anonymization purposes. Comparison of the plots demonstrates good linearity, with the slopes close to the diagonal, black line (theoretical slope = 1), and good agreement between the three transitions at each concentration point. Four replicate measurements are represented at each concentration point. Transitions 1 ( $m/z$  747.3 → 863.4), 2 ( $m/z$  747.3 → 964.5) and 3 ( $m/z$  747.3 → 1092.5) are represented by a red diamond, blue asterisk, and green triangle, respectively. In some cases, the data overlay such that transition one is not visible. The mean slope calculation across all labs in this example is 1.05 with an inter-laboratory CV of 10.3%.

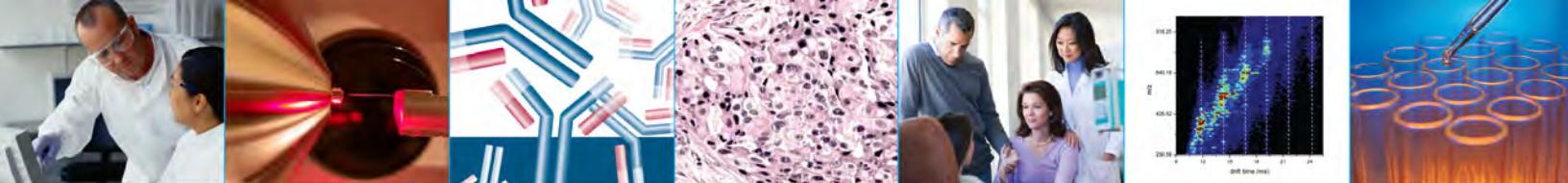




**Table 1. Target proteins and their signature peptides**

*Amino acid in red indicates stable-isotopically labeled residue. Cysteines (*underlined*) are carbamidomethylated.*

Protein	Abbrev.	Species	Signature Peptide	MH+ (mono)	MRM Transitions			
					Q1	Q3		
Aprotinin	APR-AGL	Bovine	AGLC <u>Q</u> TFVYGGCR	1493.7	747.3	863.4	964.5	1092.5
Leptin	LEP-IND	Mouse	INDISHTQSVSA <u>K</u>	1407.3	469.9	590.8	647.8	728.4
Myoglobin	MYO-LFT	Horse	LFTGHPETLE <u>K</u>	1279.7	427.2	510.3	583.8	724.4
Myelin Basic Protein	MBP-HGF	Bovine	HGFLPR	732.4	366.7	391.3	538.3	595.4
Myelin Basic Protein	MBP-YLA	Bovine	YLASASTMDHAR	1328.6	443.5	491.2	526.8	823.4
Prostate Specific Antigen	PSA-IVG	Human	<u>I</u> VGGWE <u>C</u> EK	1082.5	541.7	808.3	865.4	969.4
Prostate Specific Antigen	PSA-LSE	Human Horse	LSEPAELTDAV <u>K</u>	1280.7	640.8	783.4	854.5	951.2
Peroxidase	HRP-SSD	radish	SSDLVALSGGHTFG <u>K</u>	1483.8	495.3	711.4	798.4	982.5
C-Reactive Protein	CRP-ESD	Human	ESDTSYVSL <u>K</u>	1136.6	568.8	617.4	704.4	805.4
C-Reactive Protein	CRP-GYS	Human	GYSIFSYAT <u>K</u>	1144.6	572.8	724.4	837.5	924.5
C-Reactive Protein	CRP-YEV	Human	YEVQGEVFTKPQLWP	1826.9	914.0	1053.5	1181.6	1525.8



**Table 2. Linear Regression Parameters for Studies I, II, and III**

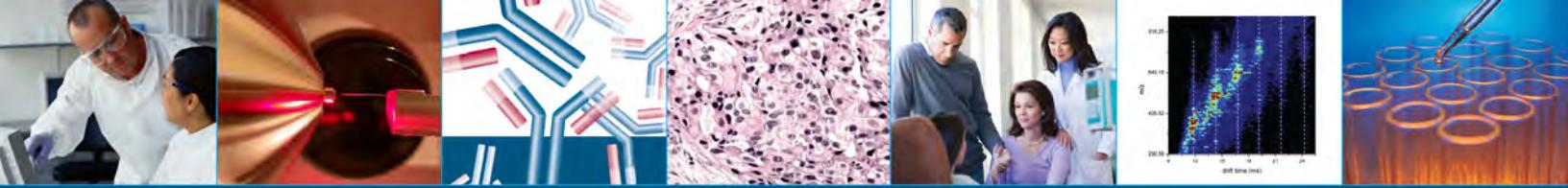
Average *a*, *b*, *c* process replicates for each peptide averaged across sites for linear slope, its inter-site CV, the median intra-site CV, the % recovery from the linear regression, and its inter-site CV. The CVs increase in distribution between studies I, II, and III, for both parameters, as expected with the increasing complexity of the three studies.

Signature Peptide	Study I					Study II					Study III <sup>a</sup>				
	Linear Slope	Lin. Slope CV	Intra-site CV <sup>b</sup>	% Recov. <sup>c</sup>	% Recov. CV	Linear Slope	Lin. Slope CV	Intra-site CV <sup>b</sup>	% Recov. <sup>c</sup>	% Recov. CV	Linear Slope	Lin. Slope CV	Intra-site CV <sup>b</sup>	% Recov. <sup>c</sup>	% Recov. CV
APR-AGL	1.047	10.3%	5.9%	104.7	10.3%	0.559	12.7%	5.5%	55.9	12.7%	0.761	31.5%	10.9%	76.1	31.5%
CRP-ESD	1.002	8.1%	4.4%	100.2	8.1%	0.570	11.7%	5.8%	57.0	11.7%	0.443	18.4%	11.0%	44.3	18.4%
CRP-GYS	1.209	8.1%	3.7%	120.9	8.1%	0.536	10.4%	3.8%	53.6	10.4%	0.166	30.1%	13.5%	16.6	30.1%
HRP-SSD	1.072	15.2%	5.8%	107.2	15.2%	0.785	18.3%	5.3%	78.5	18.3%	0.437	19.7%	16.5%	43.7	19.7%
LEP-IND	1.049	8.8%	4.9%	104.9	8.8%	0.152	10.5%	14.5%	15.2	10.5%	0.255	14.3%	25.4%	25.5	14.3%
MBP-HGF	1.100	12.6%	3.6%	110.0	12.6%	0.821	14.7%	5.9%	82.1	14.7%	0.276	24.2%	15%	27.6	24.2%
MBP-YLA	1.135	6.5%	4.2%	113.5	6.5%	0.796	4.7%	5.4%	79.6	4.7%	N/A	N/A	N/A%	0.0	N/A
MYO-LFT	1.371	16.5%	3.2%	137.1	16.5%	0.984	11.0%	3.6%	98.4	11.0%	0.565	17.3%	15.4%	56.5	17.3%
PSA-IVG	1.520	11.0%	3.9%	152.0	11.0%	0.843	10.9%	4.4%	84.3	10.9%	0.596	17.9%	12.6%	59.6	17.9%
PSA-LSE	1.137	8.4%	3.3%	113.7	8.4%	1.524	8.8%	3.6%	152.4	8.8%	0.940	10.7%	9.3%	94.0	10.7%

<sup>a</sup> Results in table are the averages across process replicates a, b, c performed in Study III, except for Intra-site CV, which is represented as the median value.

<sup>b</sup> Intra-site CV was calculated from all replicates for each peptide using a single transition. The intra-site CV represented here is the median value across all sites for each peptide by Study.

<sup>c</sup> % recovery was estimated by the slope of the linear regression.



## Next Steps

A follow-up consortium-wide study is in the planning stages, which will focus on clinical utility. These efforts will include the following:

- Extending the MRM assay dynamic range at the low end of the concentration scale.
- Increasing the number of analytes and employing scheduled MRM assays as a means to test the feasibility of expanding the multiplexing capabilities of the current protocol.
- Incorporating cancer-relevant proteins as analytes.
- Employing a more diverse set of triple quadrupole mass spectrometers to better gauge variability in the results, if any, due to distinct instrument design.

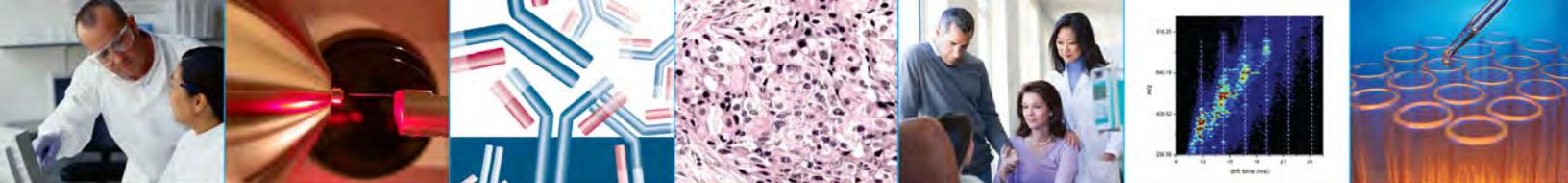
### 2.2.2 A Reference Yeast Proteome for LC-MS/MS Performance

Individual laboratories often use their own mixture of proteins/peptides to serve as a quality control sample for measuring optimal performance of liquid chromatography-tandem mass spectrometry (LC-MS/MS) platforms. The lack of a widely available complex standard, however, has hindered benchmarking of platform performance across the proteomics community. Providing a well-characterized reference proteome as a resource for the community will facilitate standardization of both current and emerging proteomic technologies, and will ensure the highest quality data are being generated.

The yeast *Saccharomyces cerevisiae* provides an attractive reference proteome for the following reasons:

- The proteome is complex, consisting of ~4,500 proteins expressed under normal growth conditions.
- The abundance of proteins ranges from fewer than 50 to more than  $10^6$  molecules per cell, which is sufficient to challenge the dynamic range of conventional mass spectrometers.
- Yeast remains the most extensively characterized complex biological proteome.
- The yeast proteome is inexpensive and easy to produce in large quantities.

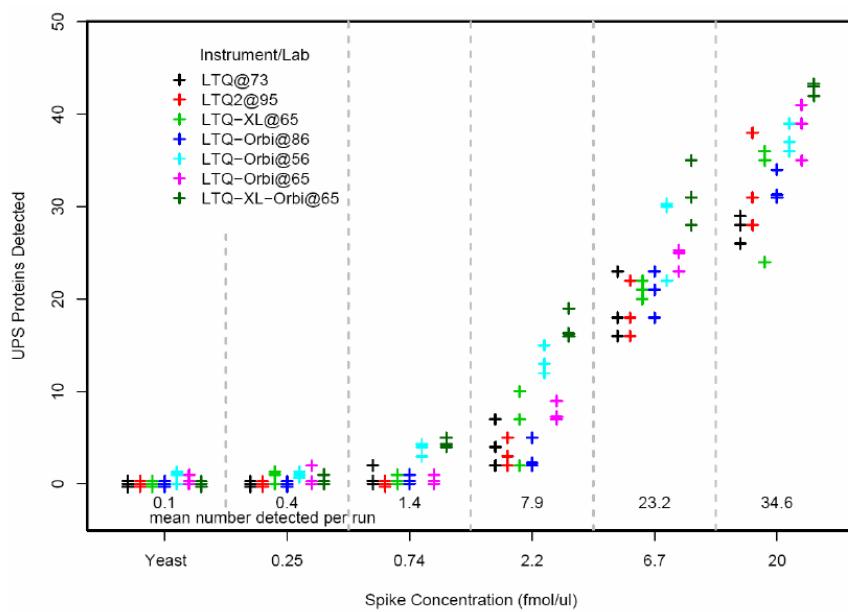
A CPTAC consortium-wide study has described a large-scale production of the yeast *Saccharomyces cerevisiae* reference proteome, which is now offered to the community through NIST. A historic reference dataset has been created which characterizes the yeast reference proteome and defines performance of ion trap-based LC-MS platforms using a series of performance metrics. These performance metrics can also be used for the differential diagnosis of underlying causes of LC-MS platform performance issues. Finally, the yeast reference proteome, when spiked with human proteins, can be used to benchmark the power of shotgun platforms to detect biomarker candidates at different levels of concentration in a complex matrix.



**A Reference Yeast Proteome for Benchmarking LC-MS Platform Performance.**  
*This manuscript was submitted to Molecular & Cellular Proteomics in May 2009.*

**Figure 12. Ion suppression and detection efficiency for Universal Proteomics Standard (UPS, containing 48 human proteins, 5 pmols of each) proteins spiked at different concentrations into the yeast matrix**

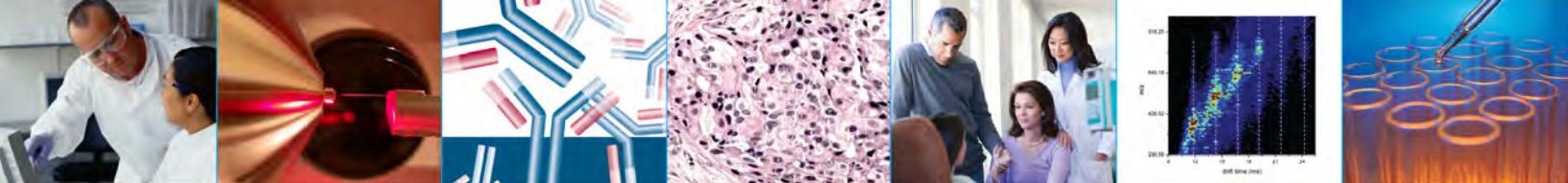
The result of each RPLC run is indicated by a “+” plotting symbol; colors denote different instruments. Plotting symbols have been jittered to avoid overplotting of identical values. Protein detection is defined as observing two or more peptides mapping to the same protein (in a single RPLC run). The number of detected UPS proteins increases with increasing spike concentration. Further, instruments exhibit heterogeneous sensitivities to increasing concentration. Note that at an equimolar spike concentration of 2.2 fmol/ $\mu$ l all instruments detect at least one UPS protein in each run.



### Next Steps

A follow-up consortium-wide study is being implemented. These efforts include:

- Comparing the performance of different shotgun proteomics analysis platforms using the yeast reference proteome, incorporating multidimensional separations of proteins/peptides.
- Detection of protein expression differences to distinguish cancer-relevant phenotypes using the Her2+ breast cancer cell line, BT474, as a model.



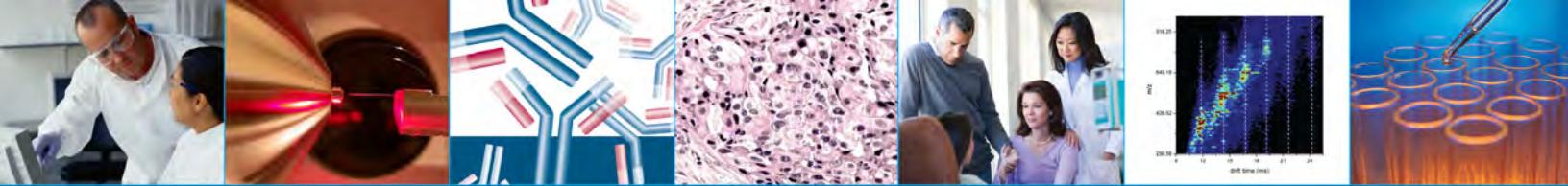
### 2.2.3 Performance Metrics for Evaluating LC-MS/MS Systems

LC-MS/MS remains the principle technology platform for the global characterization of complex proteomes, known as shotgun proteomics. Shotgun proteomics is increasingly being used to discover candidate biomarkers for cancer and other diseases by comparing the differences between normal and disease proteomes. However, this approach assumes that shotgun proteomics is sufficiently reproducible and that observed differences accurately reflect the underlying biology of disease phenotypes. A lack of objective criteria to assess the analytical system performance of shotgun proteomics makes this assumption difficult to assess.

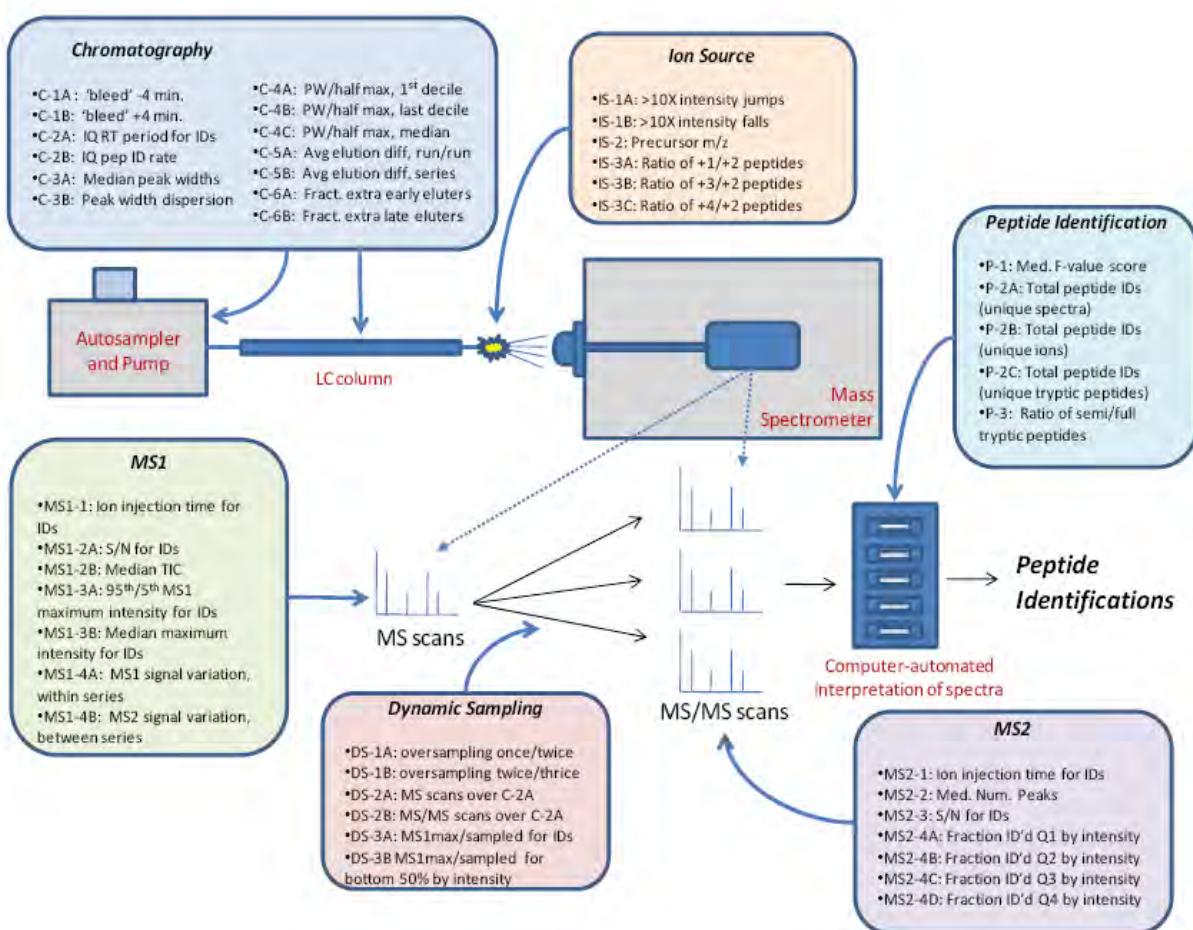
The most commonly used measure of variability in LC-MS/MS proteomic analyses is the number of confident peptide identifications. While these numbers can indicate repeatability, they cannot indicate whether system performance is optimal or which components require optimization. A CPTAC consortium-wide study has described 44 metrics for evaluating the performance of LC-MS/MS system components, and has implemented a freely available software pipeline that generates these metrics directly from LC-MS/MS data files. These metrics can characterize sources of variability in proteomic platforms, both for replicate analyses on a single instrument as well as across laboratories.

***Performance Metrics for Evaluating Liquid Chromatography-Tandem Mass Spectrometry Systems in Proteomic Analyses.***

***This manuscript was submitted to Molecular & Cellular Proteomics in April 2009.***



**Figure 13. Schematic representation of performance metrics mapped to LC-MS/MS system elements**

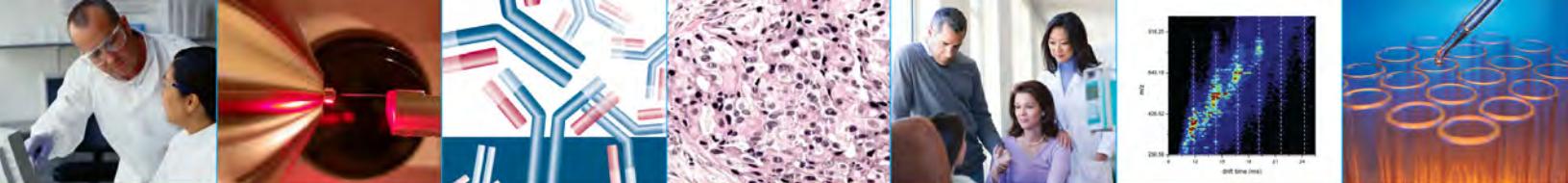


### Next Steps

A follow-up consortium-wide study is being implemented. This includes continuing development and evaluation of new performance metrics as well as distribution of software to the proteomics community which will facilitate implementation.

#### 2.2.4 Repeatability and Reproducibility in LC-MS/MS Analysis

Repeatability and reproducibility both give rise to the variability observed in analytical proteomics, which is problematic. Repeatability of the technique can be tested by measuring the variation in results from run-to-run when using a particular sample the same way on the same instrument. Reproducibility, on the other hand, describes the variation in results between two different instruments in the same laboratory or two

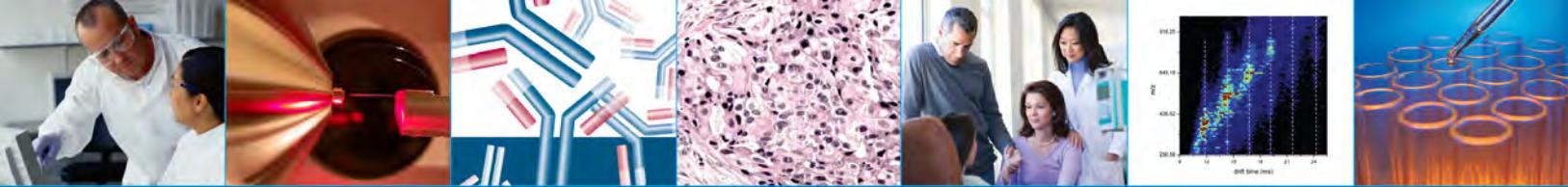


instruments in different laboratories. Repeatability and reproducibility of proteomics platforms have never been systematically evaluated. Thus, the field remains unaware of the limits of repeatability and reproducibility in LC-MS/MS proteomics and the degree to which standardization can be achieved.

The CPTAC network was created in part to address the need for standardized studies to evaluate technology platforms. To evaluate measures of variability, seven institutions contributed data from twelve instruments for three different samples, each of which differed in protein number and dynamic range. It was observed that peptide identification was associated with low repeatability and reproducibility because different sets of peptides are discovered with each run rather than the same core set of peptides. However, the values for protein identification are much greater because different subsets of peptides can be detected for each protein. This data supports a stochastic identification model rather than one in which a core set of peptides is routinely identified.

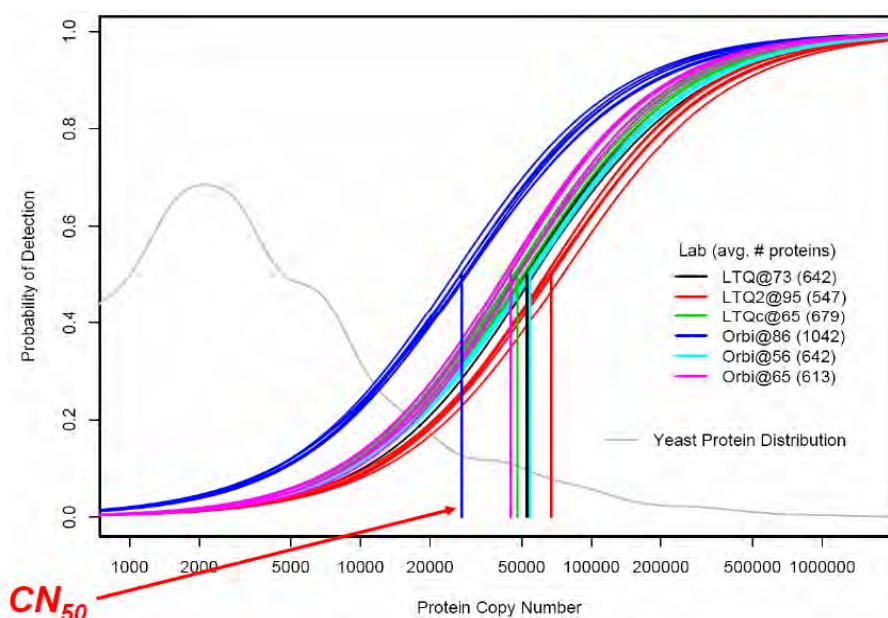
Implementation of a SOP resulted in a reduced number of peptide identifications. This suggests that diverse protocols for LC-MS/MS systems enable the identification of a greater number of peptides. Thus, in studies where the goal is to maximize discovery of peptides, an SOP would not necessarily serve a purpose as long as acceptable standards for identification are followed. For biomarker discovery, on the other hand, the goal is to carefully compare proteomes (e.g., normal and cancer), arguing that a more targeted approach is needed. Using SOP-driven methods, LC-MS/MS proteomics can distinguish proteomic differences amounting to less than 10% of the detectable proteome. Careful adherence to SOP-driven methods can effectively standardize LC-MS/MS proteomics across laboratories, optimizing biomarker discovery.

***Reproducibility and Repeatability in CPTAC Consortium Proteomic Analyses by Liquid Chromatography—Tandem Mass Spectrometry.***  
***This manuscript was submitted to the Journal of Proteome Research in May 2009.***



**Figure 14. Inter- and intra-laboratory variability in protein detection efficiency in the yeast standard**

Logistic regression curves indicate the probability of protein detection as a function of ( $\log_{10}$ ) copy number for each run. The vertical lines indicate the mean copy number (for each instrument) corresponding to 50% probability of detection ( $PD_{50}$ ). The gray curve denotes the copy number distribution of proteins in the yeast standard [PubMedID: 14562106]. The graph indicates that, on average, only the most abundant yeast proteins have high probability of being detected. Indeed, only 9% of yeast proteins have copy numbers greater than 20000.



### Next Steps

A follow-up consortium-wide study will continue analyzing repeatability and reproducibility in subsequent LC-MS/MS performances.

## 2.3 Mass Spectrometry Platforms

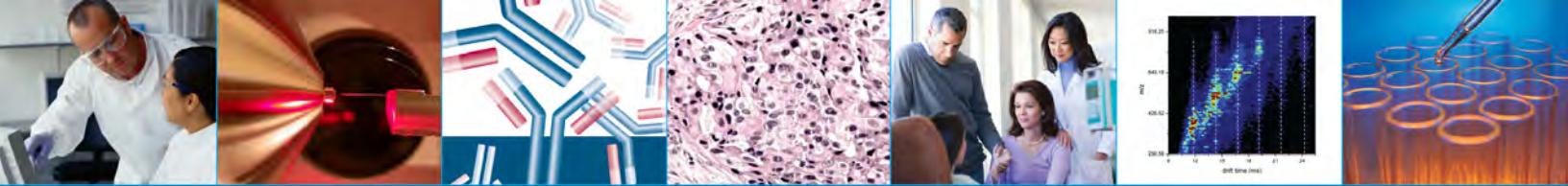
### 2.3.1 Existing and Emerging Discovery Platforms

#### 2.3.1.1 Shotgun Proteomics Technology Platforms for Unbiased Discovery

**Principal Investigator:** Daniel Liebler, Ph.D., Vanderbilt University School of Medicine

#### Research Interest:

- Extensive comparison of shotgun-based techniques
- Reverse phase protein arrays and antibody arrays



### Project Goals and Significance:

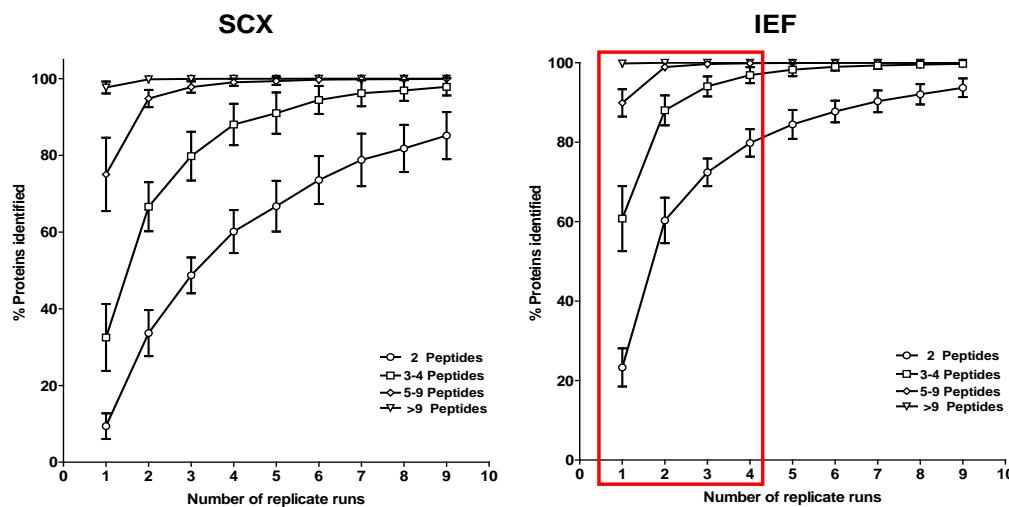
The major focus areas of the Vanderbilt CPTAC program are 1) the optimization of shotgun proteomics technology platforms for unbiased discovery of biomarker candidates in tissues and proximal fluids, 2) the development and optimization of targeted mass spectrometry-based assays for biomarker candidates in tissues and plasma and 3) the development and standardization of data analysis tools for analysis of proteomics data.

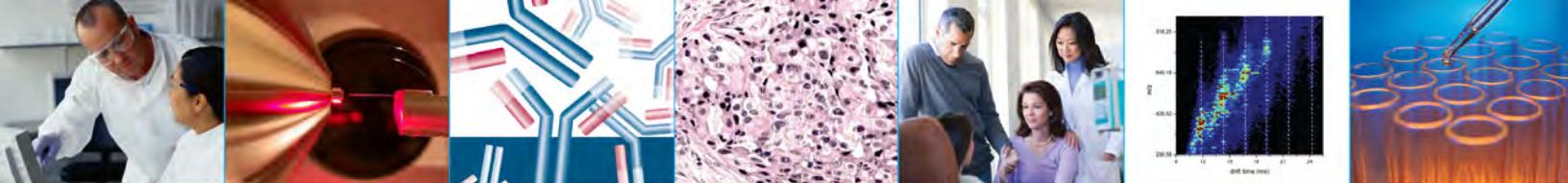
### Accomplishments:

**Validation of isoelectric focusing (IEF)-based peptide fractionation as superior to strong cation-exchange (SCX) for multidimensional LC-MS/MS shotgun proteomics (Slebos et al (2008)).** Shotgun proteomics has been widely used to analyze complex proteomes for biochemistry and cell biology, but not to systematically detect proteomic differences between disease-specific phenotypes. This is due mainly to variability in the proteome inventories obtained with the most widely used shotgun proteomics platform, which combines SCX peptide fractionation with reverse phase LC-MS/MS analysis. Recent reports suggesting that IEF peptide separations offered superior performance led us to compare the efficiency and reproducibility of the two approaches. Analysis of a colorectal adenocarcinoma tryptic digest on both SCX and IEF platforms demonstrated that IEF fractionated the peptides much more efficiently and that the IEF-based platform resulted in convergence on >90% of protein identifications by 3-4 peptides within four analytical replicates, whereas the SCX platform required eight replicates to reach 90% of possible identifications (Figure 15).

**Figure 15. Cumulative protein identifications for SCX and IEF-based shotgun proteomics platforms**

*Proteins identified by 2 peptides, 3-4 peptides, 5-8 peptides and >9 peptides are shown as a percentage of all proteins identified by each platform. The IEF platform identified over 90% of medium abundance proteins (identified by 3-4 peptides) in 4 replicates, whereas the same fraction of identification with SCX required 6 replicates.*

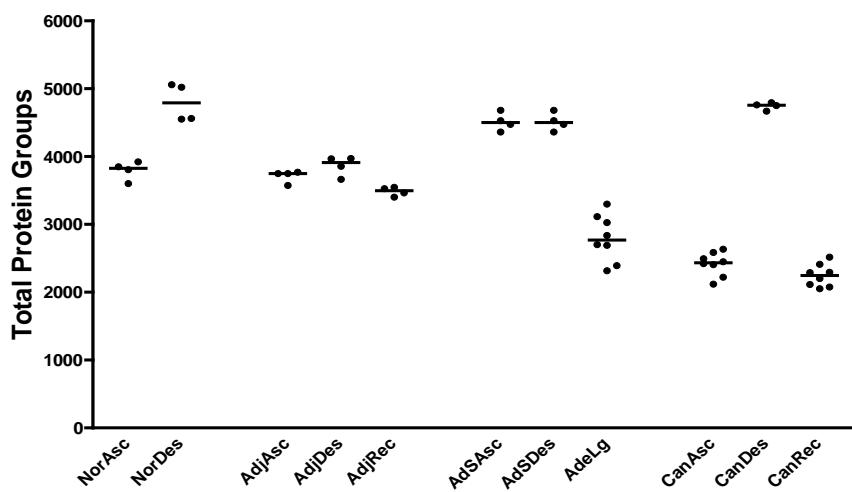




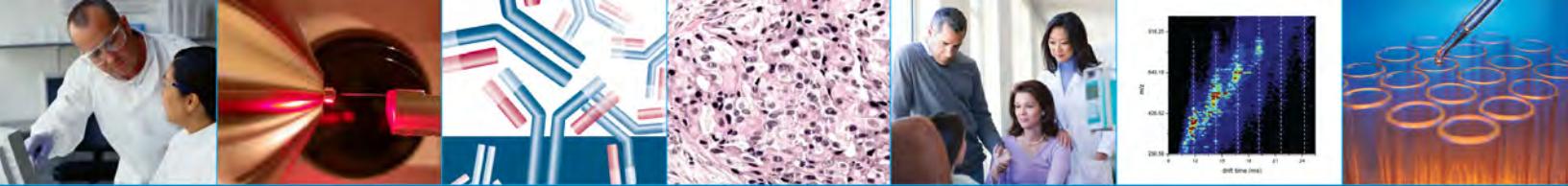
**Applied a standardized shotgun proteomics platform to compare proteomes of human colon, colorectal cancer and precancer tissue specimens.** The IEF-LC-MS/MS platform was applied in a large survey of proteomes from colon adenocarcinomas, large tubulovillous adenomas, small tubular adenomas and normal colon epithelium to evaluate proteotypes of colon precancers and cancers. Four replicate analyses of pooled proteome samples from each tissue type from ascending and descending colon and rectum demonstrate high reproducibility of protein identifications between replicates (Figure 16). Shotgun proteome inventories of each tissue type generated 3,000 to 4,000 proteins with confident identifications (at least two peptides per protein). These are the largest human tissue proteome inventories acquired to date and are approximately an order of magnitude beyond what has been achieved with previous generation platforms (e.g., 2D gels).

**Figure 16. Protein identifications from pools (n=20) of colon tissue specimens using IEF-based shotgun proteomics platforms**

*Specific tissue samples were normal ascending (NorAsc) and descending (NorDes) colon epithelium; apparently normal tissue adjacent to resected adenocarcinoma from ascending (AdjAsc) and descending (AdjDes); large (>1 cm) tubulovillous adenomas from all locations (AdeLg); small (< 1 cm) tubular adenomas from ascending (AdSAsc) and descending (AdSDes) colon; and adenocarcinoma (Stages 1-3) from ascending (CanAsc) and descending (CanDes) colon and rectum (CanRec). Mean protein inventories (horizontal bars) were approximately 3,500 confidently identified proteins and CVs for replicate measurements (symbols) were <10%.*



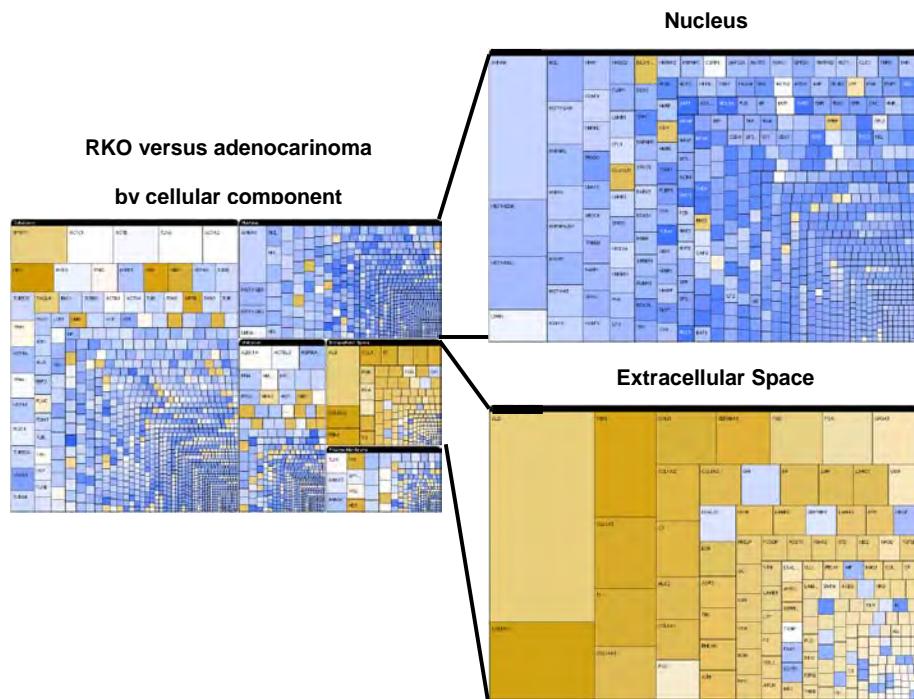
**Developed a statistical model for distinguishing proteotypes corresponding to phenotype differences from shotgun proteomics datasets.** To deal with the over/under dispersion observed in shotgun data, the team applied a quasi-likelihood Poisson model that has no restriction on distribution assumptions. To correct for



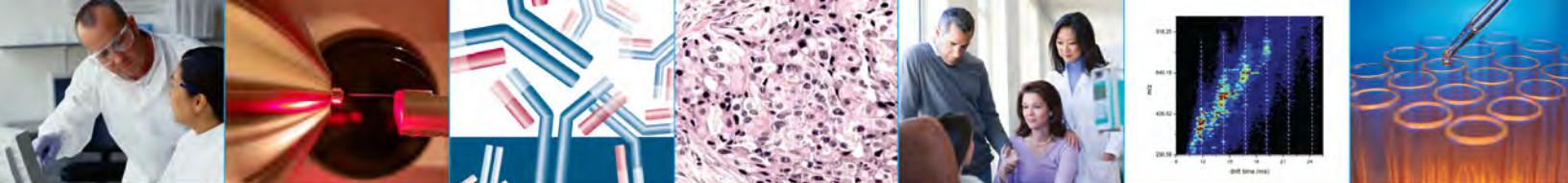
potential selection bias in multiple comparisons of protein count data, the team applied the False Discovery Rate (FDR) method. A treeview visualization was generated using the HoneyComb package (Hivegroup, Richardson, TX), which displays all identified proteins as squares in a larger field, classified by cellular component using Ingenuity Pathways (Ingenuity Systems, Redwood City, CA) (Figure 17). The size of each square represents the total number of spectral counts in the full dataset and the color represents the ratio expression for each protein between the datasets being compared. This visualization tool enables cancer biologists to directly inspect and interpret shotgun proteome inventories of tissue biospecimens and to discern proteomic differences at the level of biology. Tools to present complex proteomics datasets to cancer biologists will be critical to the process of prioritization of biomarker candidates.

**Figure 17. Treeview visualization of shotgun proteome inventory comparison between a rectal adenocarcinoma and a colorectal cancer cell line (RKO)**

*The color reflects the rate ratio as calculated by the statistical model with proteins that were observed with higher spectral counts in RKO in blue and proteins observed at higher spectral counts in the adenocarcinoma in brown.*



**Evaluated shotgun proteomic analysis of formalin-fixed, paraffin-embedded (FFPE) tissues.** The team adapted IEF-LC-MS/MS shotgun proteomics platform to the analysis of FFPE tissues and demonstrated that FFPE tissues yield proteome inventories equivalent to fresh frozen tissues. Moreover, FFPE tissues stored up to 10



years under ambient conditions yield proteome inventories equivalent to those of fresh tissues. This work demonstrated the feasibility of performing retrospective biomarker discovery studies on archived tissue specimens where clinical outcomes are known in advance and where bias can be minimized by careful selection of archival specimens in accord with accompanying clinical information (Sprung et al. (2009)).

**Developed and implemented an integrated, open-source data analysis pipeline for shotgun proteomics.** This system combines the Myrimatch database search algorithm with the IDPicker utility for data filtration and parsimonious protein assembly under conditions of fixed false discovery rates. These tools exceed the performance of other available tools and the integrated pipeline has become the data analysis pipeline for CPTAC interlaboratory studies involving shotgun proteomics. This work was supported both by a CPTC R01 award (to Dr. D. L. Tabb) and the Vanderbilt CPTAC program (Tabb et al. (2007); Zhang et al. (2007)).

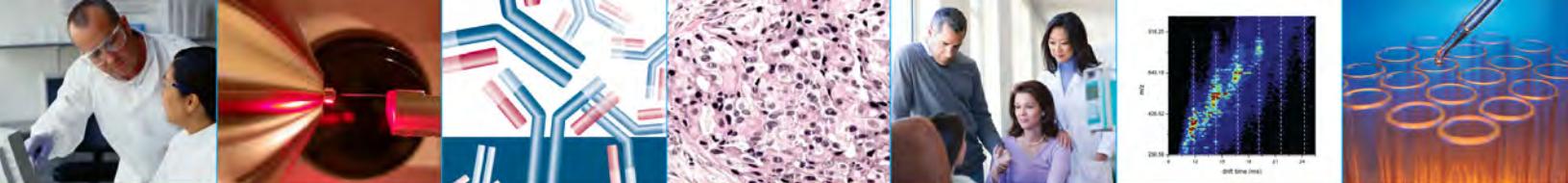
**Developed a “label-free” method for targeted quantitation of proteins.** The team has demonstrated that targeted liquid chromatography-multiple reaction monitoring mass spectrometry (LC-MRM-MS) of tryptic peptides enables quantitation of biomarker candidate proteins without the need for labeled peptide standards. Measurements by this method are consistent with western blotting, but display lower variation and do not require antibodies. This approach provides an effective means to rapidly configure specific, targeted assays for biomarker candidates (Liu et al. (2009), submitted).

#### Next steps:

- Development of standardized methods for quantitative analysis of phosphoproteins.
- Development of a comprehensive informatics pipeline for targeted quantitative analysis of proteins by LC-MRM-MS.
- Development of a standardized method for label-free quantitation of protein biomarker candidates in tissue specimens.
- Development and standardization of a hybrid immunoaffinity-LC-MRM-MS method for biomarker candidate analysis in plasma.
- Extension of quantitative LC-MRM-MS methods to FFPE tissue specimens.
- Identification of quality metrics for plasma specimens for proteomic analysis.

#### Collaborations:

**Development of a comprehensive informatics pipeline for targeted quantitative analysis of proteins by LC-MRM-MS (with David Tabb (Vanderbilt) and Michael MacCoss (Washington)).** The application of LC-MRM-MS to targeted quantitation of proteins (via analysis of component peptides) is hindered by a lack of experimental design and data analysis tools to handle highly multiplexed analyses of hundreds of peptides and to organize and analyze data in hierarchical designs needed for biomarker verification experiments. The Vanderbilt CPTAC established a subcontract in October 2007 with Dr. MacCoss to coordinate with related efforts begun at Vanderbilt by Dr. Tabb. The system component in development in the MacCoss laboratory, called Skyline, facilitates the *in silico* prediction of proteotypic peptides, peptide transitions, and



chromatographic retention time. This work is being pursued in consultation with CPTAC collaborators at the Broad Institute (Dr. Steve Carr) and the Fred Hutchinson Cancer Research Center (Dr. Mandy Paulovich). Both Skyline and Myrmidon will be implemented in test versions in mid-late 2009. These tools will be made available as open-source applications to the research community.

**Development of MS/MS spectral libraries for peptide identification (collaboration with Drs. Steve Stein and Paul Rudnick at the National Institute for Standards and Technology).** Peptide sequences are most commonly identified from MS/MS spectra by database searching, which although very effective is error prone and computationally intensive. An alternative approach is library matching, where the spectrum of an unknown is searched against a library of spectra from defined compounds. The group has collaborated with Drs. Stein and Rudnick to provide them with a collection of MS/MS spectra derived from shotgun proteomic analysis of colon, lung, oral cavity and stomach normal and tumor tissues. The initial transfer of data to NIST occurred in December 2008 and consisted of 47,141 different sequences, which resulted in an expansion of the existing NIST spectral library by 27%. A second transfer of a similar set of data is planned in January 2009.

### 2.3.1.2 Assessment of Serum Peptide Profiling to Detect Cancer Specific Patterns

**Principal Investigator:** Paul Tempest, Ph.D., Memorial Sloan-Kettering Cancer Center (MSKCC)

**Research Interests:**

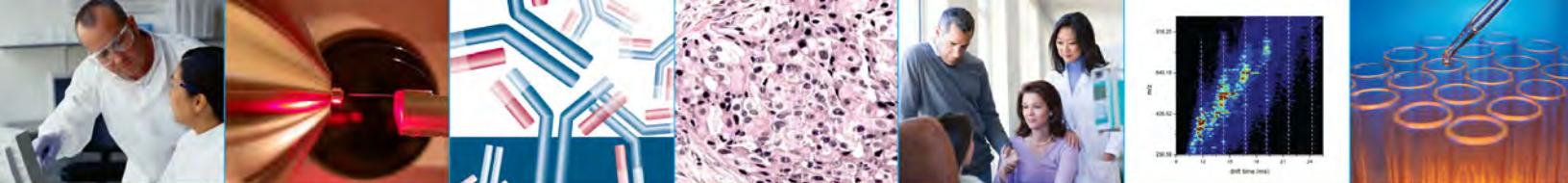
- Sample fractionation using magnetic beads for capture of peptides
- Automated sample processing technology (robotics)

**Project Goals and Significance:**

As cancer involves transformation and proliferation of altered cell types that produce high levels of specific proteins and enzymes, such as proteases, it will not only modify the array of existing blood proteins ('proteome') but also their metabolic products, i.e. peptides ('peptidome'). The MSKCC team has proposed to evaluate and document whether serum peptide patterns, or the protease activities producing them, can be measured reproducibly and whether they have diagnostic value for cancer detection, or mark a given clinical outcome, or distinguish clinically insignificant from significant cancer.

**Accomplishments:**

**Major improvements on reproducibility and repeatability for MALDI-TOF were achieved.** Initial comparisons of MALDI-TOF reproducibility between the MSK prototype set-up and the NYU mirror-site for serum peptide profiling indicated good concordance in numbers of detected peptides but significant variations in MALDI-TOF ion-intensities, even after numerous efforts to fine-tune the instruments towards similar outputs. The

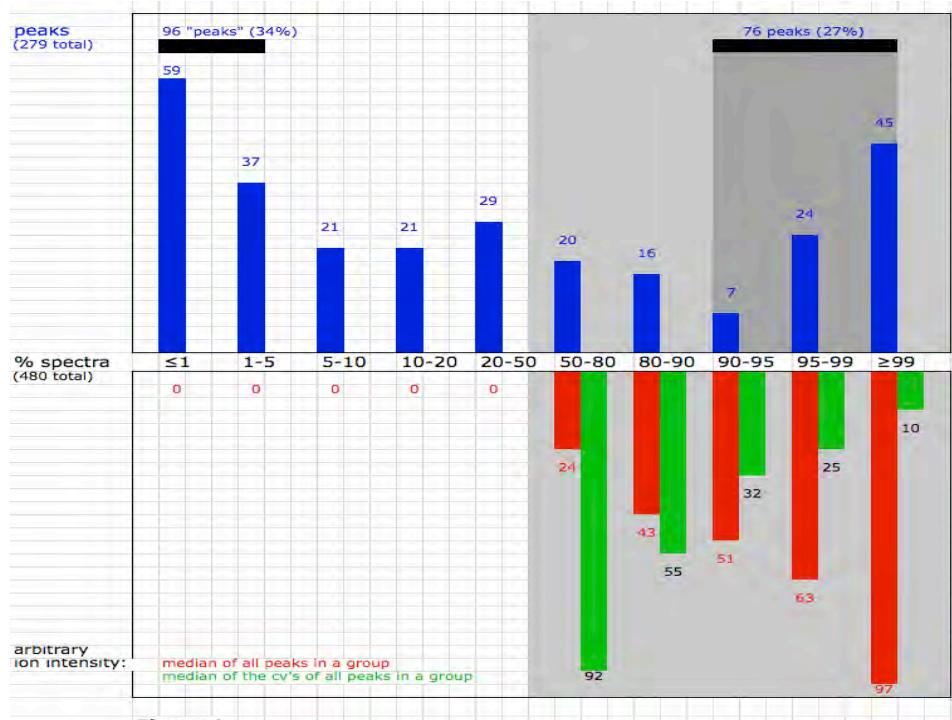


team has since made major improvements on (i) aligning peaks along the x-axis ( $m/z$ ) and (ii) normalizing ion intensities (y-axis).

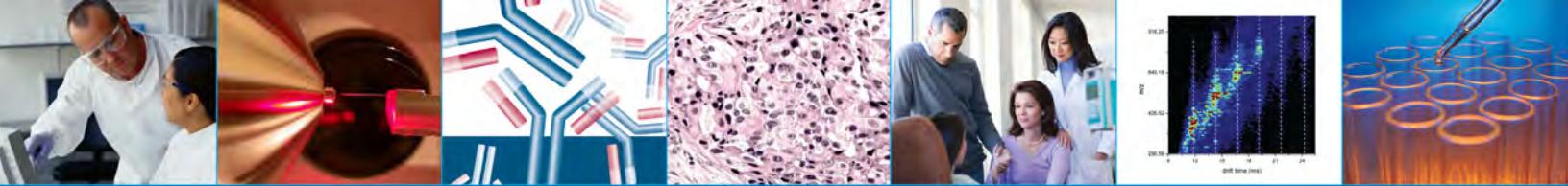
Three studies on repeatability and reproducibility, all using aliquots from a single reference serum sample, have been completed: (1) sample spotted (“written”) by robot A and analyzed (“read”) by MALDI-TOF instruments A or B; (2) sample written by robots A or B, and read by MALDI-TOF A; (3) samples are completely separately written / read by robot/MALDI combinations A/A and B/B. The total number of observed unique peaks (in 480 replicates) was 279, about 110-160 per spectrum. Peaks varied in ion intensity (v.i. = 0.1-250) and reoccurrence (1/480 to 480/480) between replicates (Figure 1). 76 peaks were present in  $\geq 99\%$  of all replicates, but another 96 were only observed in  $\leq 5\%$  of the replicates (blue bars in Figure 18). Well-represented peaks tend to have higher ion intensities (red bars) and lower ion intensity CV's (green bars). Box plots for the ion intensity CV's of the top 10% most represented peaks within each site (A or B) or across two sites (overall) for all three studies are shown in Figure 19.

Sample ‘writing’ (study 2) induces more overall variability than ‘reading’ (Study 1) (14% versus 11%) and the overall median CV. (sample aliquots separately ‘written/read’ at different sites in study 3) is 17%. Again, CV’s are bigger for small, less frequently ( $\leq 90\%$ ) occurring peaks, and smaller for the tall, highly frequently ( $\geq 99\%$ ) ones. These observations put practical limitations on the selection of diagnostic peak patterns.

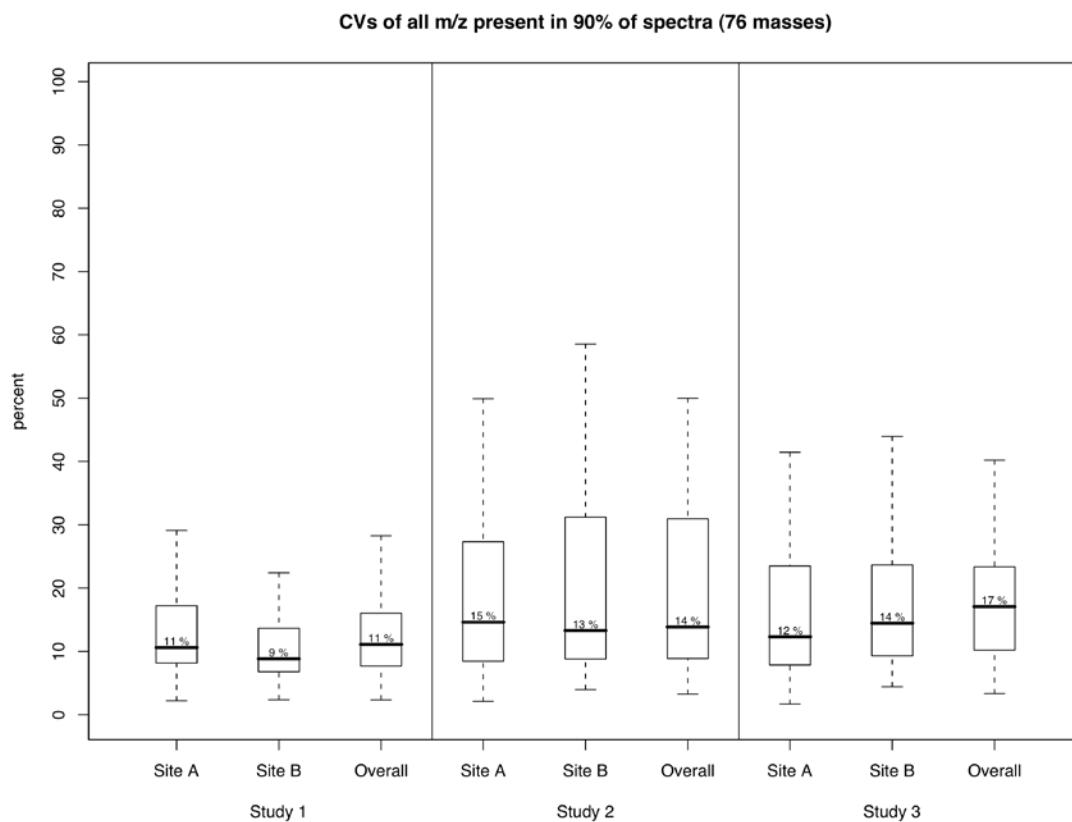
**Figure 18. Reproducibility and repeatability of MALDI-TOF**



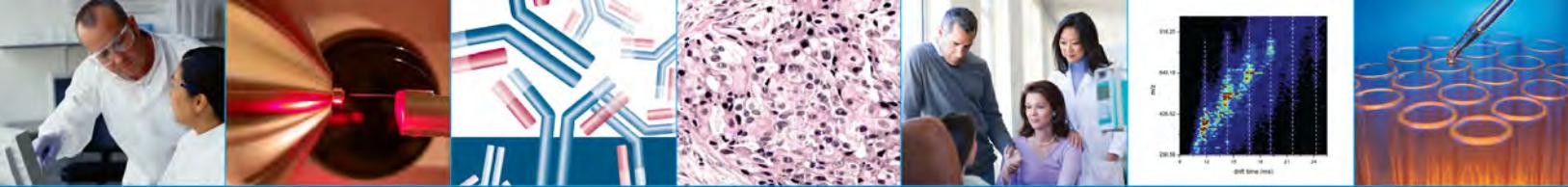
**Figure 1**



**Figure 19. Reproducibility and repeatability of MALDI-TOF**

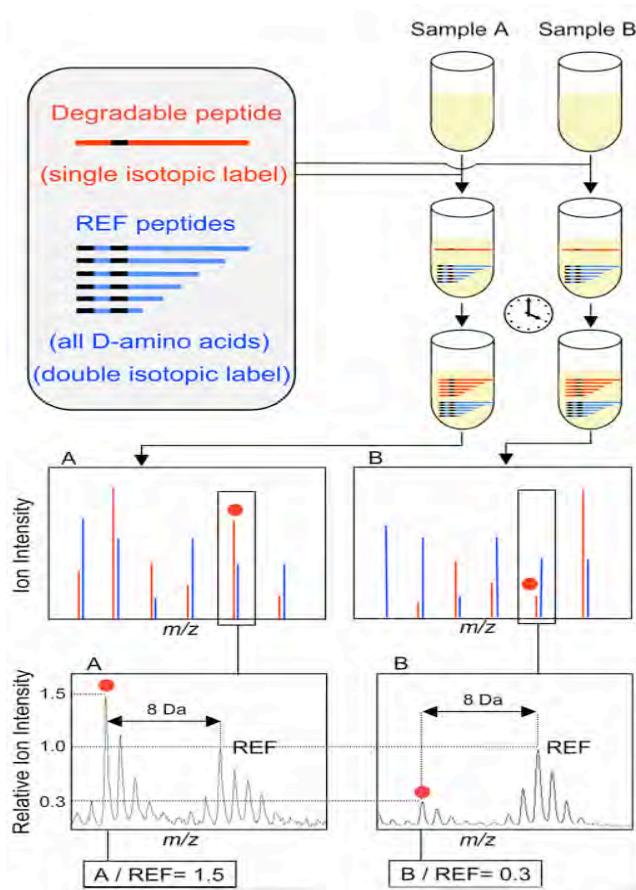


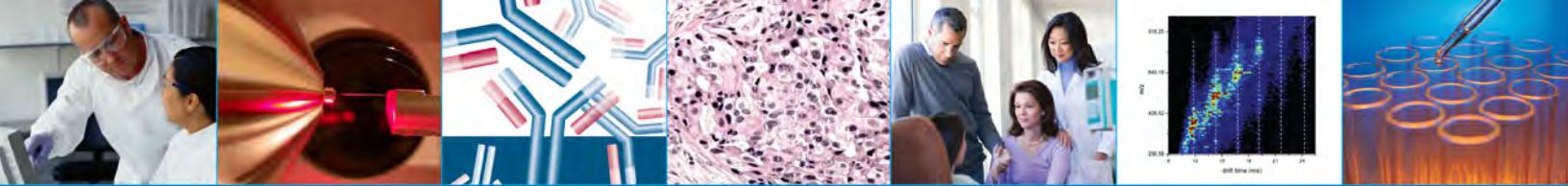
**Developed an exopeptidase activity-test.** The team introduced a test of the global exopeptidase activities within individual proteomes of two or more groups of biological fluids, by accurate quantification of degradation products and multivariate statistical analysis of the resulting qualitative and quantitative patterns. Figure 20 depicts the schematic of a test that satisfies all the above criteria.



**Figure 20. Strategy for monitoring differential exoprotease activities in biological samples**

A degradable, isotopically labeled substrate and its corresponding set of non-degradable, double-labeled reference peptides (representing a nested set of truncated sequences) are added to two samples A and B that may, for instance, represent two different clinical states. At a given time point, the two reactions are stopped and analyzed by magnetic-based solid-phase extraction and a MALDI-TOF mass spectrometric read-out. Spectra are processed, ion intensities normalized versus the total ion current, matching peptide pairs assigned, and the ratios of each of the degradation products versus the corresponding spiked reference peptides calculated. This ratio represents the relative ion intensity of each peptide that is comparable across all samples.





This test tracks degradation of artificial substrates, under strictly controlled conditions, using semi-automated MALDI-TOF mass spectrometric analysis of the resulting patterns. Each fragment is quantitated by comparison with double-labeled, non-degradable internal standards (all-D-amino acid peptides), spiked into the samples at the same time as the substrates to reflect adsorptive and processing-related losses. The full array of metabolites is then quantitated (CVs of 6.3 to 14.3% over 5 replicates) and subjected to multivariate statistical analysis. We analyzed serum samples of 48 metastatic thyroid cancer patients and 48 healthy controls, with selected peptide substrates taken from earlier standard peptidomic screens (i.e., the 'discovery' phase), and obtained class predictions with 94% sensitivity and 90% specificity, without prior feature selection (24 features).

Advantages of this approach over standard peptidome measurements are robustness (samples yielded the same results after 4 hours on the bench as did fresh samples); reproducibility (nicely timed and otherwise controlled); and quantitation.

**Reference materials generated.** Peptide substrates produced and tested in serum exoprotease assays. A number of non-degradable reference peptides have been made and available to date including fragments corresponding to FPA, C3f, clusterin, and A1AT.

#### Collaborations:

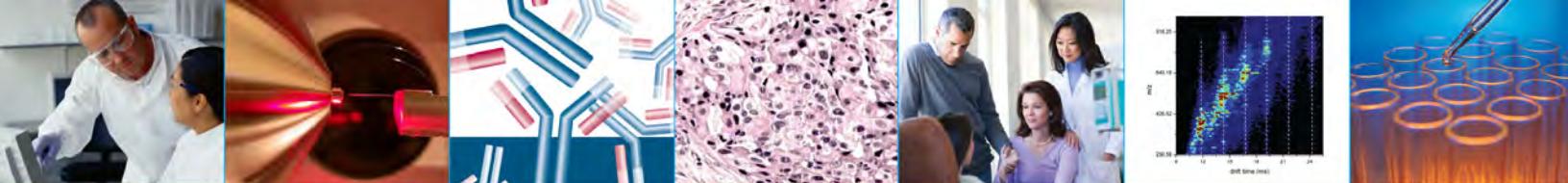
##### ***Mark Robson and Clifford Hudis, Breast Cancer Service, MSKCC***

The specific aims of the Breast Team project are: 1) to collect serum and plasma samples from women undergoing breast biopsy for radiographic or physical exam abnormalities and, 2) to compare the serum peptide profiles of women whose biopsies do or do not result in a malignant diagnosis. The objective is to identify specific peptide patterns associated with a malignant diagnosis.

The protocol for the Breast Team project (IRB#06-003) is designed to limit sources of clinical variability that could confound interpretation of the laboratory output. Participants are all unaffected by previous cancer, and samples are drawn before the actual biopsy itself. These parameters were chosen to limit confounding by possible occult malignancy (or effects of treatment) or by biased sampling based upon knowledge of pathology results. As of January 8, 2009, 230 patients have been consented and we have pathology results for 225; 51 had cancer (16 pure DCIS and 35 invasive).

##### ***Hans Lilja and James Eastham, Prostate Cancer Service, MSKCC***

Protocol 1 (# 08-114): assess the value of serum peptide profiling and exopeptidase assays to predict biopsy outcome in men undergoing prostate biopsy. We collaborate with New York Presbyterian Hospital (Cornell), New York, where 200 men ( $\geq 95\%$  of Caucasian/ European American background) undergo prostate biopsy each year, and at SUNY Downstate Medical Center, Brooklyn, which services a large, multi-ethnic population of men comprised of  $\geq 90\%$  men of African descent. The study protocol (#08-114) was approved by the MSKCC IRB on October 16, 2008. Amended IRB-protocols incorporating all required changes have been submitted for approval at



the Cornell IRB office on December 23, 2008 and at the SUNY Downstate IRB office on January 7, 2009.

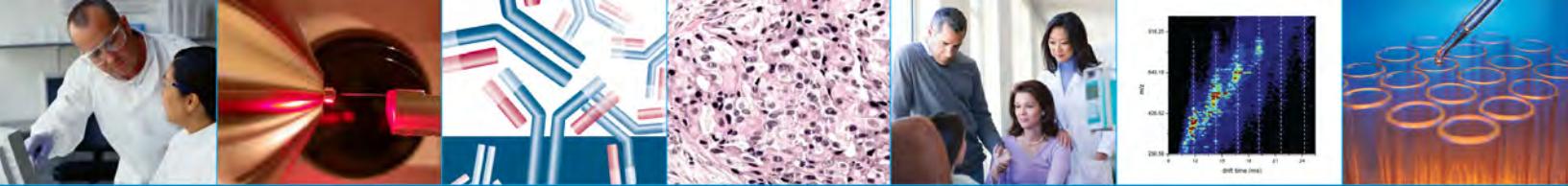
Protocol 2 (# 04-132): The team will determine whether serum peptide profiling or exopeptidase assays may be used to characterize patient risk in men undergoing radical prostatectomy. As of January 8, 2009, 222 patients have been enrolled in this protocol. The team has pre-operative blood sample from 191 patients, and for 149 patients the team also have matching postoperative blood samples.

#### **Brett Carver, HOPP, MSKCC**

To explore whether exopeptidase assays show differences between mice with and without prostate cancer, control mice and genetically altered mice with cancer were analyzed. All mice displayed normal germ line genotype with prostate specific genetic deletions of *Pten* and *p53* to avoid off target background effects. The genetically altered mice were analyzed along a spectrum of prostate tumorigenesis at 6 months of age. Selected differences in degradation patterns were observed after incubation with sera from wildtype (WT) mice or with those with cancer. An overlap with the degradation patterns (cancer vs. control) in human sera was found. The most striking differences in the peptide-degradation patterns were metabolites C3f-12 and C3f-15 (on average 2-to 3-fold higher corrected-ion-intensities following incubation of the substrate in sera of mice with cancer as compared to sera from WT controls) and Clus-12 (3-fold higher corrected-ion-intensity after incubation in ‘cancer’ versus ‘control’ sera).

#### **Next steps:**

- The standard (robotics/MALDI-TOF) peptide profiling repeatability and reproducibility studies will be concluded and a manuscript prepared.
- The team will continue to accrue patient samples.
- Exopeptidase assays. The team will continue to investigate and optimize degradation conditions and times for each of the synthetic substrate peptides (existing and possible novel ones) in each of the sera / plasmas from the different groups of cancer patients and controls. The team will also determine permissible inter-mixability of the different substrates, and particularly of their resulting degradation ladders, in order to avoid disturbing the peak patterns (by ion suppression effects) and to avoid overlapping isotopic envelopes (when the peaks are too close).



### 2.3.1.3 Targeted and Global Proteomic Strategies for Early Breast Cancer Detection\*

\*Funding for this Center began in Year 2 of the CPTC Initiative

**Principal Investigator:** Susan Fisher, Ph.D., University of California, San Francisco

#### Research Interests:

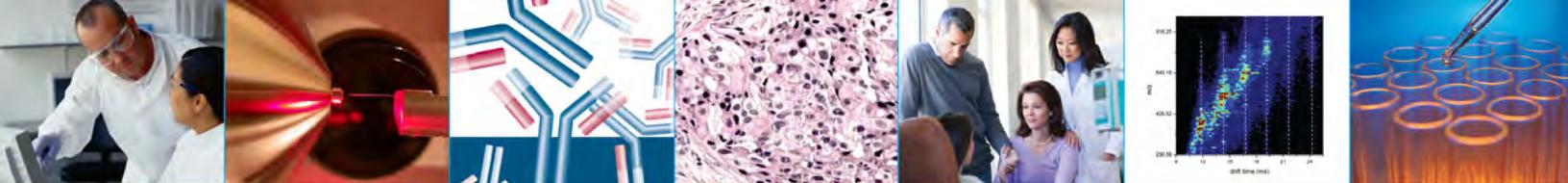
- Characterization of post-translational modifications (PTMs)
- Development of <sup>18</sup>O-based technique to monitor degree of degradation of samples (protease activity monitoring)

#### Project Goals and Significance:

The UCSF team proposed to evaluate proteomic technologies that will enable the early detection of several tumor types through the application of blood-based tests. The group, consisting of scientists at LBNL, UCSF, The Buck Institute, MD Anderson Hospital, and the University of British Columbia, has the broad expertise that this project requires. The team has special interest in breast cancer. Initially, the team proposed to examine both global strategies and targeted MS-based approaches to develop optimal workflows for the identification of protein signatures of human breast cancer cells. The global strategies will utilize multiple workflows that emphasize quantitative comparisons. The targeted approaches will focus on cancer-specific proteins that result from aberrant RNA splicing. Candidate biomarkers will be validated using reverse phase protein arrays. The requisite antibodies will be generated by Epitomics. The next phase of the project will employ human clinical samples. Specifically, the team will apply an analogous, optimized approach for analyzing plasma samples that will be prospectively collected from breast cancer patients ( $N = 200$ ). Control samples will be obtained from healthy women with benign breast disease and from women with rheumatoid arthritis to account for the contribution of proteins associated with inflammatory processes. The candidate approach targets both the spliceome, which will be profiled using breast cancer biopsies from the plasma donors, and posttranslational modifications (e.g., glycosylation, phosphorylation, proteolysis, and oxidative damage). The team will also include a plan for establishing a systematic way to standardize proteomic protocols and data analysis among the groups that exploits curability to analyze mass spectra generated on numerous platforms and the robust statistical methods that will be employed to mine these large data sets. In the end, the group envisions that in conjunction with the other CPTAC teams and the NCI, they will develop methods, tools, and reference samples for the research community that will make the promise of MS-based cancer biomarker discovery a reality.

#### Accomplishments:

***The team developed the SOP and evaluated the performance of discovery platforms based on LC-MALDI instruments (AB4800 and Thermo-Fisher vMALDI LTQ) using low (NCI20) and high complexity (yeast digest) samples.*** The MALDI platforms were unique within the consortium and were not included in the mainstream Unbiased Discovery workgroup (WG) study. Likewise, the newly acquired QStar-Elite at the Buck Institute was also evaluated as an LC-ESI discovery platform by employing a



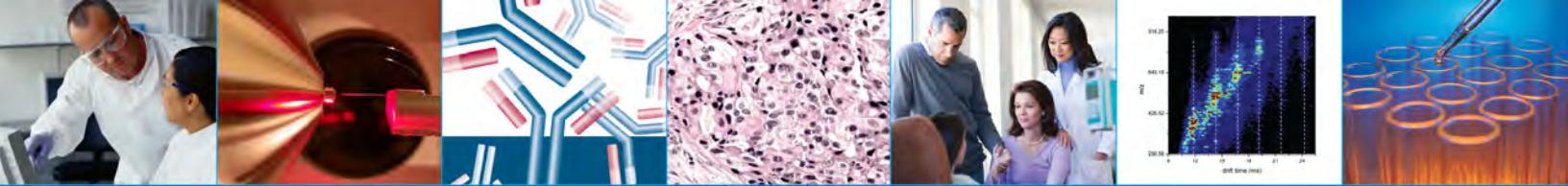
customized LTQ/Orbitrap SOP. The data from these studies will be publicly available on the Tranche website.

Both LC-MALDI platforms generated similar numbers of peptide and protein identifications in the unbiased discovery study. The AB4800 instrument equipped with more advanced acquisition software offered higher time efficiency. ESI LTQ/Orbitrap instruments outperformed both LC-MALDI platforms in the number of identifications made and in the speed of data acquisition. Generation of LC-MALDI data prompted expansion of the consortium's repertoire of software tools that are used to efficiently handle MALDI-derived raw data. In contrast to LC-MALDI platforms, the performance of the QStar-Elite platform was equivalent to that of the LTQ/Orbitrap platforms.

***Considerable progress has been made in the development and implementation of affinity-based workflows for detecting proteins that carry various cancer-related PTMs.*** These include glycoproteins, phosphoproteins and proteins that carry modifications that are indicative of oxidative damage. The affiliated team, Buck Institute, has been developing methods to prepare protein standards with 3-nitrotyrosine and 4-hydroxynonenal modifications. For glycoprotein analyses, UCSF and Buck Institute investigators have assembled a set of commercially available glycoproteins from Sigma consisting of fetuin,  $\alpha$ 2-hs-glycoprotein, follicle stimulation hormone, lactoferrin, ribonuclease B glycoprotein, and yeast invertase. A subset has been incorporated into PTM WG glycoproteomics workflow as either positive or negative controls.

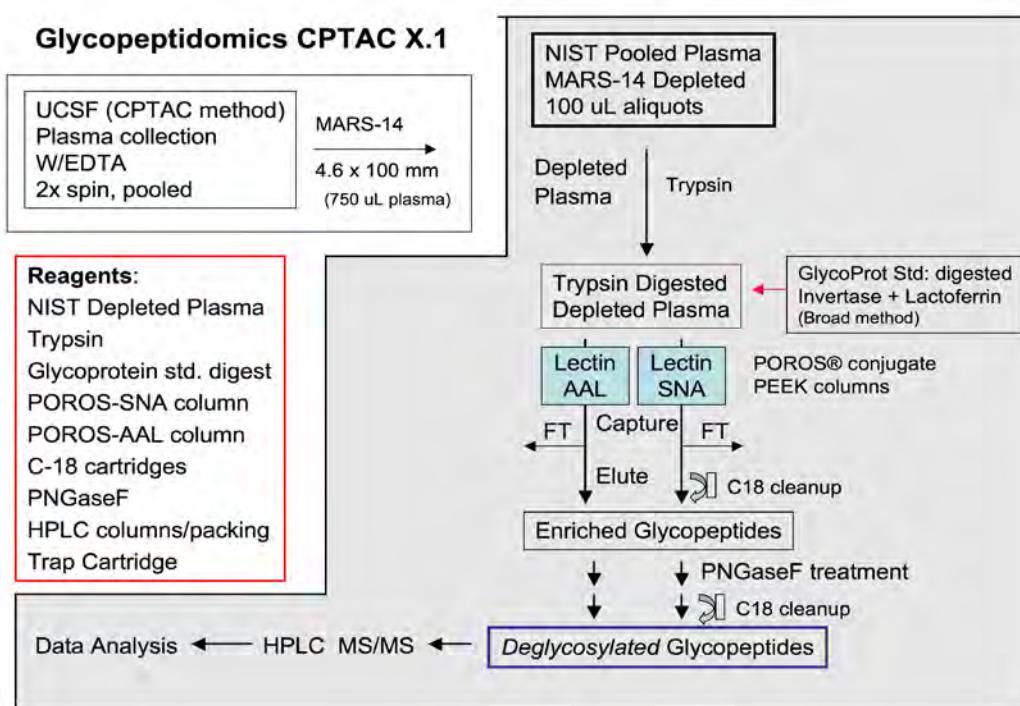
To date, workflows for glycoproteins and phosphoproteins have been the major focus, and the team is close to launching a large inter-laboratory CPTAC experiment targeting glycopeptides. A MS-based workflow was designed that incorporates chromatography on affinity matrices formed from lectins, proteins that bind specific glycan structures (Figure 21). Protocol optimization also included determining the appropriate ratio of starting material to column capacity, identifying the most efficient capture and elution buffers, and monitoring the PNGase F-treatment to ensure full deglycosylation. Future directions include using this workflow to perform mass spectrometry-based discovery experiments on plasma from breast cancer patients and control individuals.

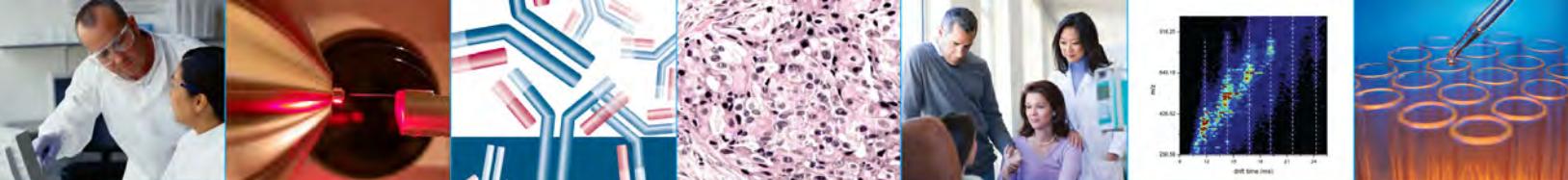
***Identified alternatively spliced breast cancer transcripts that can be used for early detection.*** The group is focusing on markers that are specific for metastasis-prone (basal) subtypes that are likely to have the largest impact on survivorship. UCSF investigators are using this same cell resource to identify cell surface glycoconjugates that are markers of the same subtype. The team Co-Investigator, Dr. Gordon Mills (M.D. Anderson Cancer Research Center), is optimizing reverse phase protein array platforms for validating the results of MS-based analyses. LBNL investigators found approximately 1700 genes that showed strong evidence of alternative splicing across all the breast cancer cell lines. Then they identified the splice variants that were highly expressed in the basal subtype: CD44, FLNB, CLTC, PLEC1, FAT, FER1L3, DST, SLK, MYO6, and FAM62B. In complementary efforts, UCSF investigators showed that basal subtype cell lines and actual breast cancers express the sulfated high-affinity L-selectin carbohydrate ligand defined by the glycan structures that are recognized by the MECA-79 antibody



(Figure 22). Interestingly, CD 44 is one of the glycoproteins that carry this very unusual carbohydrate structure. Dr. Mills's reverse phase protein array platforms have now been optimized for the detection of potential biomarkers using less than 1nl of patient (pancreatic cancer) and control plasma.

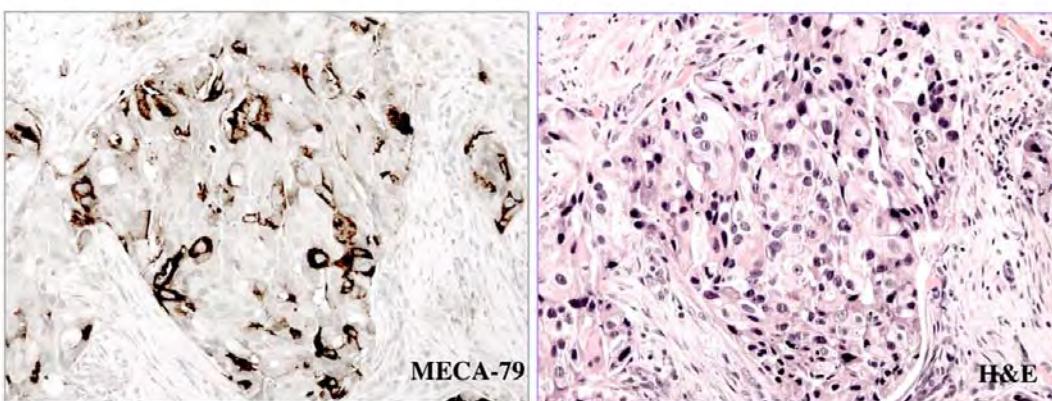
**Figure 21. Schematic of the proposed plasma glycopeptidomics inter-laboratory CPTAC study**





**Figure 22. The MECA-79 epitope**

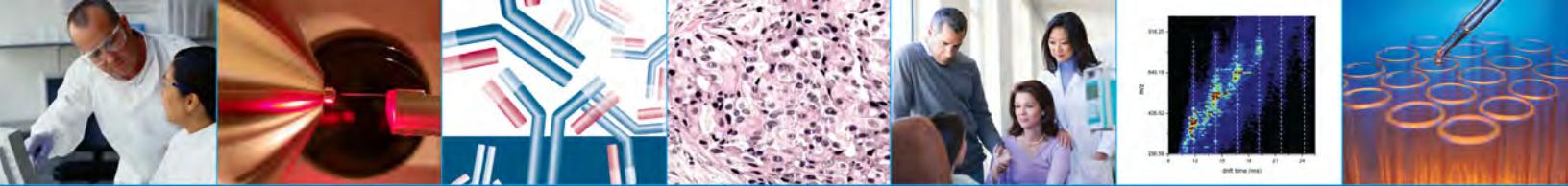
The MECA-79 epitope, a sulfated carbohydrate ligand for L-selectin (which mediates leukocyte rolling and tethering in the blood-vascular space), is also expressed by human basal-type breast cancer cells *in situ*. We have functional evidence that this glycan plays a role in metastasis. H&E, hematoxylin and eosin.



The Team's impact on consortium-wide endeavors has been particularly noteworthy for the progress of the inter-laboratory studies. The UCSF team has been heavily involved with the CPTAC Experimental Design and Statistics (Verification Studies Working Group) and the Bioinformatics WG. Their contributions in the verification experiments helped this group reach a major conclusion that by using common materials and SOPs, MS-based assays of proteins in plasma can be sensitive and highly reproducible across laboratories and instrument platforms. In addition, the resulting experimental protocol provides a foundation for proteomics investigators to develop similar MS-based protein assays in their own laboratories. In the Bioinformatics working group, the UCSF team facilitated the successful processing of data from the full set of Unbiased Discovery Studies, the development of an inter-laboratory study characterizing database search identification algorithms, and the design of tools to make CPTAC data sets compatible with caBIG. In PTM WG, the Buck Institute investigators have developed an approach that uses direct chemical modification of protein mixtures to produce nitrosylated, carbonylated, and HNE-modified proteins. The UCSF team played a large role in developing the protocol for biospecimen collection (e.g., processing and storage of plasma) and in selection of a web-based clinical data collection tool.

**Next steps:**

- Joint efforts between the UCSF group and the LBNL lab of Dr. Mark Biggin are underway to develop iterative MS/MS acquisition (IMMA) routines to improve the performance of LC-MALDI platforms in protein discovery applications. Likewise, the protocols for utilizing iTRAQ reagents for relative protein quantitation that were jointly developed for bacterial samples will be evaluated for their utility in plasma-based biomarker discovery studies. UCSF and Buck Institute will participate in all the inter-lab Digestion WG studies by performing the assays and analyzing the samples. The team will participate in the next round of Discovery



WG experiments that entail analysis of protein standards that are spiked into a breast cancer cell lysate that will be produced by UCSF/LBNL.

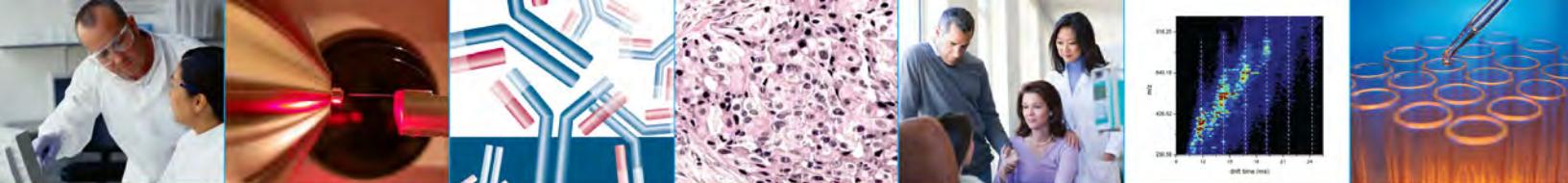
- One of the team top priorities for the new fiscal year will be to finalize a lectin-based plasma glycopeptidomics experiment as a major inter-site CPTAC study in 2009. The team also plans to carry out a large discovery-based glycopeptidomics experiment in plasma from cancer and control patients to increase our depth of analysis. If cancer-specific glycopeptides are identified in these studies, the group will move on to a verification study using a MRM experimental format.
- The team plans to work with Dr. Ron Beavis at the University of British Columbia, as well as members of CPTAC, to address the lack of integrated informatics tools in PTM-based biomarker discovery. Specifically, the team proposes to develop a MS/MS spectral database specifically for PTM-modified peptides and to link this tool to various resources such as the PhosphoSite® and GlycoSuiteDB online databases that have key information regarding experimental and/or predicted structural PTM sites.
- The team's major goal is to translate discoveries concerning basal (breast cancer) subtype-specific alterations in alternative splicing and glycosylation into MS-based assays for quantifying levels of these isoforms in patient and control plasma samples. With regard to splice variants, the group is constructing MRM assays to detect the novel splice junctions. Once sensitivity and specificity are established, the optimized assay will be used to analyze the human plasma sample banks that the group has been collecting. With regard to cancer-specific glycoforms, the same overall strategy will be used to construct SISCAPA assays for detection of basal subtype-specific species. In parallel, reverse phase protein array strategies and Firefly isoelectric focusing approaches will be used to confirm the results of these MS-based analyses.

### **Collaborations:**

Close collaborations have been established with the Bioinformatics WG, and specifically Drs. David Tabb of Vanderbilt and Paul Rudnick of NIST. In collaboration with David Tabb (R01 CPTC grant recipient), UCSF provided data files obtained from multiple replicates of HPLC MS/MS runs, in order to develop software tools to better access cross-platform data formats for ScanSifter data conversion tools and ID Picker 2.0 protein assembly tools.

The team has worked closely with Dr. Sean Seymour of Life Technologies (formerly AB) on application of the ProteinPilot® search engine (with Paragon® algorithm) in protein discovery and relative quantitation studies.

The team has ongoing interactions with scientists at Sigma-Aldrich that are aimed at obtaining high quality glycoprotein standards that contain a diversity of *N*- and *O*-linked glycan structures. Currently, the team is using a subset of these glycoproteins in the



evaluation and optimization of team's glycopeptidomics workflows that employ SNA and AAL lectin enrichment.

The team is working with Dr. Lee Makowski at ANL on the development of <sup>15</sup>N labeled phosphoprotein standards. Currently, a set of kinases has been targeted for expression, and these will be used to treat a subset of the <sup>15</sup>N labeled proteins that have been made at ANL. Once prepared, the team proposes a full, in depth characterization by HPLC MS/MS and to use these standards in evaluating team's phospho-proteomic workflows.

The Buck Institute and UCSF groups have worked extensively with Dr. Fred Regnier, team leader of the Purdue group, to develop glycopeptide and glycoprotein affinity capture methods using immobilized lectins. The team has weekly teleconferences and yearly meetings at the Buck Institute to discuss and plan method development for PTM-based biomarker discovery.

The Bay Area team has established a strong working relationship with the Susan Love Research Foundation. As part of the "Million Woman Army" that they are recruiting to enable clinical research, they are helping the team establish another sample (plasma) collection site at Alta Bates/Summit Medical Center with sites in Berkeley and Oakland California. These valuable biospecimens will be made available to the entire consortium.

#### 2.3.1.4 A Proteomics Platform for Quantitative, Ultra-High Throughput, and Ultra-Sensitive Biomarker Discovery

**Investigator:** Richard Smith, Ph.D., Battelle Pacific Northwest Laboratories

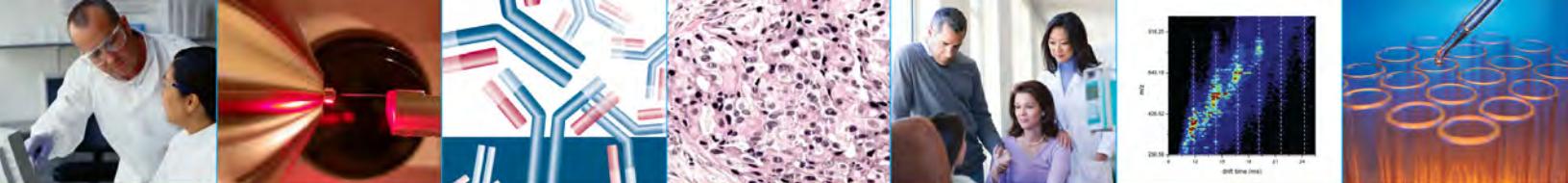
**Project Goals and Significance:**

Dr. Smith is developing a cancer biomarker discovery and validation platform for the analysis of human blood plasma and other clinically relevant samples that will provide measurements which are more robust, of higher sensitivity, higher throughput, and improved quantitative utility in comparison to current platforms. Thus, the new platform should increase dynamic range and proteome coverage to allow quantitative measurement of candidate cancer biomarker proteins present at low concentrations (i.e. PSA). The instrumental platform will be an integration of fast liquid LC based separations, a novel microfabricated array nanoelectrospray ionization (nanoESI) interface, and high resolution gas phase ion mobility separations coupled with accurate mass time-of-flight mass spectrometry.

**Accomplishments:**

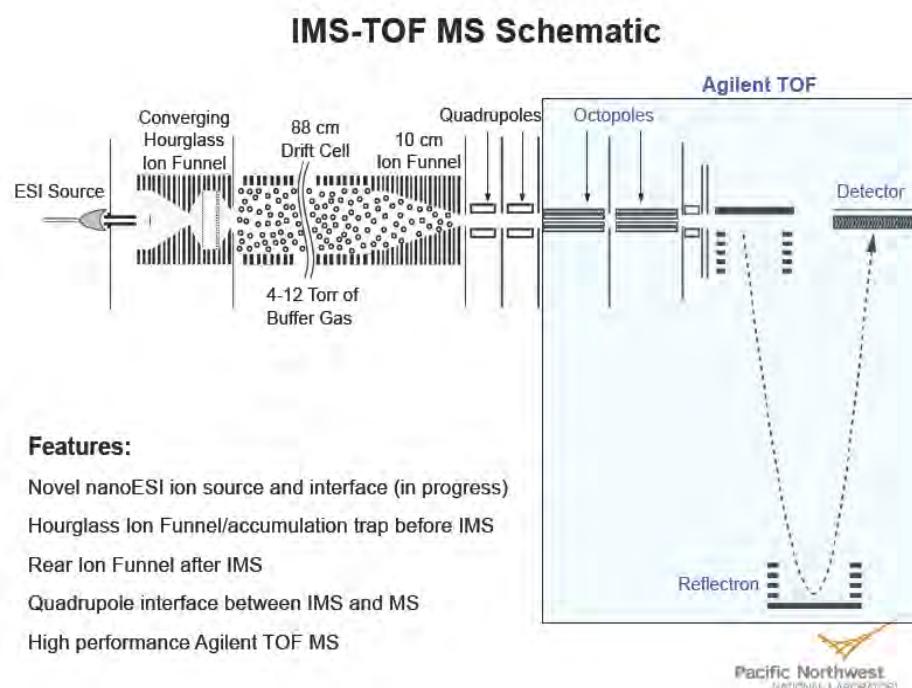
***Development of LC-nanoESI-IMS-TOF Mass Spectrometry Platform***

To achieve higher levels of throughput, sensitivity and broader dynamic range than current LC-MS techniques, a next generation proteomics platform has been developed. This new platform combines fast LC separations, ion mobility spectrometry (IMS) and time-of-flight (TOF) MS (Figure 23). To enable more sensitive measurements using the LC-IMS-TOF MS systems, an advanced multi-emitter ESI source was developed which

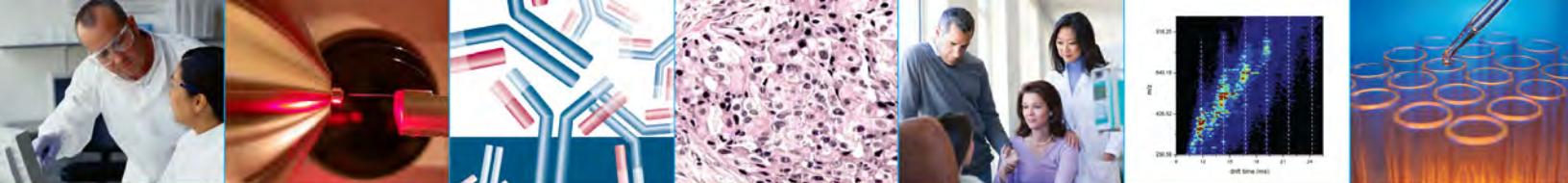


provides >10 fold increase in ion production rates relative to the conventional single ESI emitter. This fully automated four column capillary RPLC system is capable of attaining a peak capacity of ~100 in 15 minute separations. A multiplexed IMS separation approach that exhibits ~100 fold greater duty cycle than that of the conventional signal averaging IMS. It is a standalone data acquisition system that incorporates analog and digital PCI cards for gating ion packets into the IMS drift tube and synchronizes the IMS gate pulses with a TOF mass spectrometer and a 10 GHz time to digital converter card for signal detection and visualization.

**Figure 23. LC-nanoESI-IMS-TOF Mass Spectrometry Platform**



**Developed an expanded software package utilizing the Accurate Mass and Time (AMT) tag proteomics approach.** It was necessary to modify and create informatics tools to manage and process the enormous quantities of data generated from the high throughput multidimensional LC-IMS-MS analyses. Previous informatics development enabled the AMT tag proteomic approach. AMT tag proteomics is a high throughput analysis methodology for LC-MS data that requires matching elution times and accurate masses of detected features to a database. The LC-IMS-MS platform adds the additional dimension of IMS drift time to the high throughput analysis such that new and expanded capabilities to the software were required and developed. The current identification



algorithms have been expanded to include the use of drift time in the feature identification process and it is being added to the current analysis tool set.

**Validation of new LC-IMS-TOF MS platform.** A key evaluation of the utility of the platform, a complex tryptic digest of mouse plasma spiked with 20 reference peptides at varying concentrations (1 ng/ml to 10 µg/ml) was analyzed using both the traditional LC-(LTQ)-FT MS platform with a 100 minute gradient and the LC-IMS-TOF MS with a 15 minute gradient. In the LC-FT MS study, only 14 of the 20 spiked peptides with concentrations  $\geq$  100 ng/ml could be detected. In contrast, the LC-IMS-TOF MS platform was able to detect 19 of 20 spiked peptides at all concentration levels. The only peptide not detected in the LC-IMS-TOF MS analysis was ACTH fragment 18-39, which was injected at the lowest spiking level (1 ng/ml). However, fibrinopeptide A was observed at this concentration (1 ng/ml), so the 1 ng/ml spiking level appeared to be either near or below the limit of detection for the mouse plasma sample with the LC-IMS-TOF MS platform.

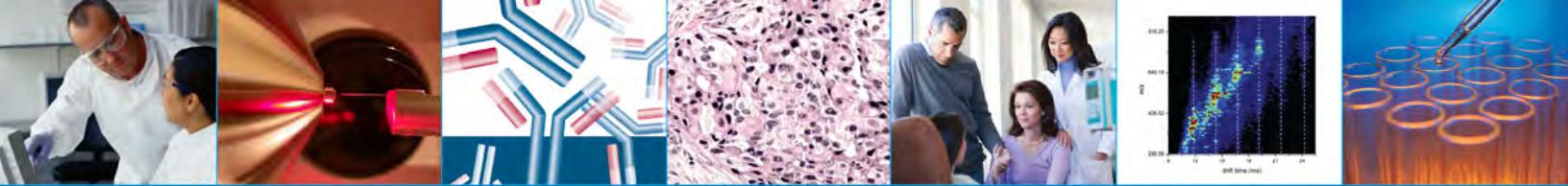
The mass spectra from both the LC-MS and LC-IMS-MS were evaluated to better understand the observed differences in detection. Close examination revealed that high concentrations of mouse plasma peptides interfere with the detection of both peptides in the LC-MS analysis. The LC-IMS-MS system removes this interference by the drift time separation capability, which allowed feature detection software to accurately characterize features that were undetectable without drift time separation. The LC-IMS-TOF MS platform achieves both sensitivity and throughput levels that are approximately an order of magnitude greater than those previously achieved in conventional LC-MS analysis.

#### **Next Steps:**

Future studies will focus on addressing key aspects of the technology in order to make it much more robust and effective. The three main challenges to be addressed are 1) establishing a high efficiency fragmentation method after the drift cell to reduce false identifications, 2) increasing the mass accuracy of the TOF MS, and 3) continued development of data management and analysis algorithms and applications.

#### **Academic/Industry/Research Collaborations:**

Agilent Technologies  
M.D. Anderson Cancer Center  
Pacific Northwest National Laboratory  
ARUP Laboratory  
GenWay Biotech Inc.



### 2.3.2 Targeted Platforms

#### 2.3.2.1 Measuring Cancer Biomarker Candidates by Targeted MS and Antibody Enrichment

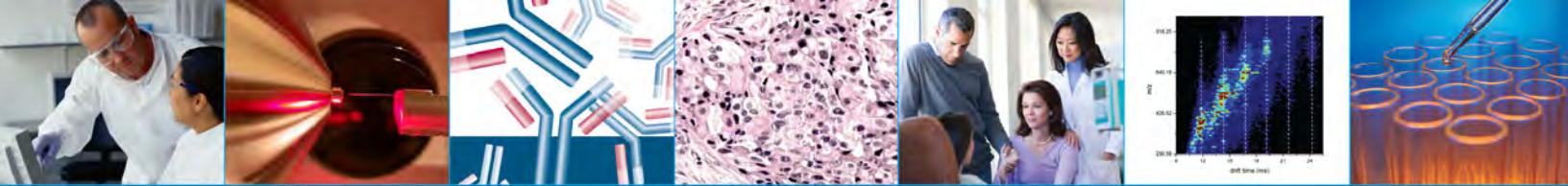
**Principal Investigators:** Steven A Carr, Ph.D., Broad Institute; Amanda Paulovich, Fred Hutchinson Cancer Research Center; Leigh Anderson, Plasma Proteome Institute

**Research Interests:**

- MRM assays for quantification
- SISCAPA

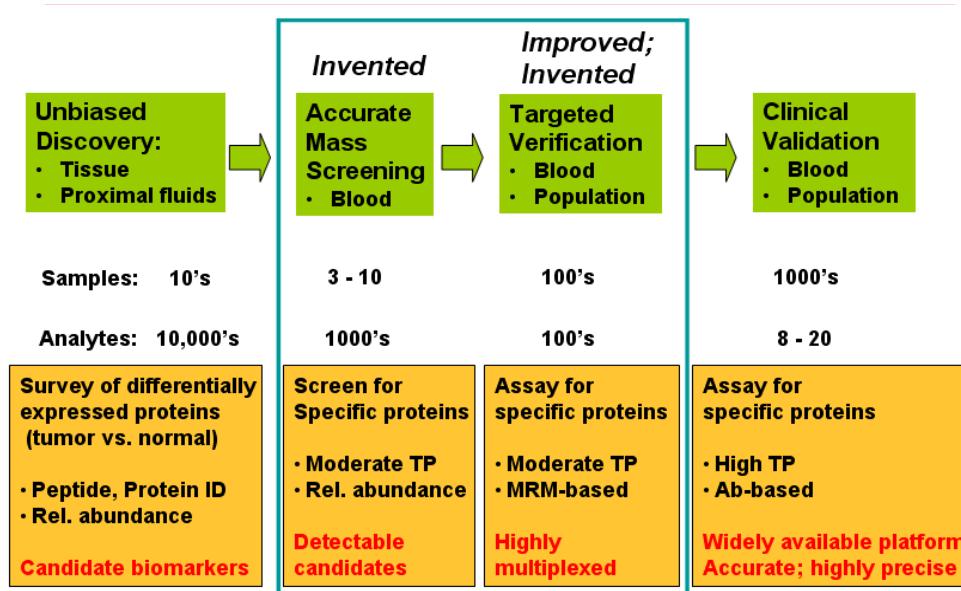
**Project Goals and Significance:**

The absence of methods capable of assessing large numbers of protein biomarker “candidates” emerging from discovery “omics” experiments is the primary bottleneck impeding development of improved cancer diagnostics. To address this issue, the Broad team led by Dr Steven Carr is interested in developing and applying novel methods based on MRM-MS. MRM-MS is a targeted quantitative approach, capable of being highly multiplexed, and that provides near-absolute structural specificity and precise quantitation when referenced to internal standards. This approach overcomes the limitations involving specificity (off-target binding and interferences), multiplexing and development costs associated with conventional ELISA assays for biomarker verification. To improve the sensitivity and specificity of MRM-MS for assaying biomarker candidates in plasma, the team is developing a range of sample fractionation and enrichment methods to enable quantitation of proteins at the low nanogram/mL (and lower) concentrations in plasma, a range where many clinically useful markers reside. A major focus of the team is to combine MRM-based methods with SISCAPA, a technology that employs peptide-specific antibodies to improve sensitivity, speed and robustness of the assays through specific analyte enrichment. SISCAPA-MRM assays will be developed for >100 candidate biomarker proteins and used to measure their levels in plasma from breast cancer patients. Outcomes will demonstrate 1) that sensitive/specific assays can be made quickly and inexpensively 2) that assays can be multiplexed, reducing the cost per analyte, and 3) that protocols, reagents and technology can be standardized and distributed.



**Figure 24. The Broad team notional pipeline for biomarker discovery through clinical validation**

*The program is focused on the items in the blue box. Figure is adapted from Rifai, Gillette and Carr. Nat. Biotech. 2006.*

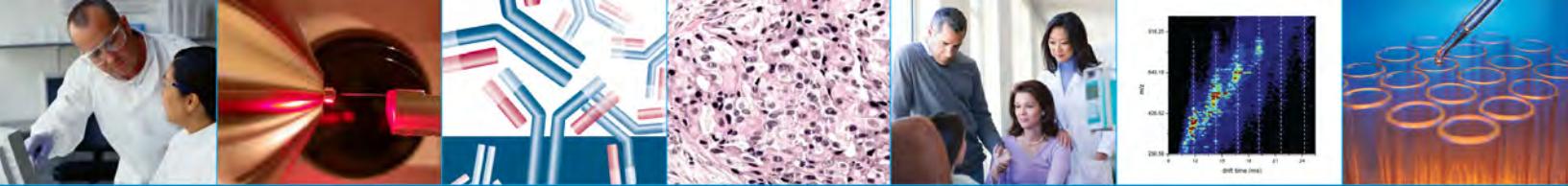


#### Accomplishments:

**Our team demonstrated, for the first time, that multiplexed, MRM-MS-based assays can be reliably constructed and robustly applied to quantify biomarker candidates at low ng/mL in plasma with CV's of 10-20%.** The team has increased the sensitivity of MRM multiplexed assays for proteins in plasma by more than 500-fold.

**Our team demonstrated, for the first time, that SISCAPA-MRM assays can quantify proteins at low ng/ml levels directly from trypsin digests of 10 to 100 µl of plasma.** Additionally the team constructed more than 20 SISCAPA-MRM assays including the first 9 and 10-plex SISCAPA assays. The sensitivity of SISCAPA-MRM assays for proteins in plasma have been increased greater than 500-fold compared to start of study. The Broad has demonstrated the real power of SISCAPA for proteins of clinical interest (e.g., osteopontin, mesothelin, IL-33 and TNF) and ability to multiplex. Several manuscripts are in preparation and one has been submitted.

**Our team established an optimized antibody reagent production and QC pipeline to support creation of SISCAPA assays for candidate cancer biomarkers.** A variety of challenging technical issues have been addressed to improve the efficiency and throughput of this process. This pipeline has been used so far to generate reagents for >100 CPTAC candidate proteins, as well as 16 other proteins targeted in four separately-funded biomarker projects.



- The Broad team has made substantial improvements in the SISCAPA methodology, optimizing several key steps of the process and implementing a common automated protocol for all team sites to use, which will allow us to fully assess the run-to-run and lab-to-lab variability of SISCAPA.
- A major improvement in assay flexibility has been achieved by migration to magnetic bead capture format in 96-well plates. Anti-peptide antibodies are added to sample digest to bind target peptides. Protein G-coated magnetic beads then capture antibodies with their bound peptide cargo. Washing and elution of the magnetic beads for MRM analysis has been automated at all three assay development sites using commercial robotics (Kingfisher).
- A further improvement in recovery of antibody bound peptides has been achieved using a novel “BeadTrap” technology, in which magnetic beads carrying Ab and target peptides are washed and eluted inside a 150 µm ID capillary forming an integral part of the analytical LC-MS/MS instrument. This automation approach (paper in press) effectively eliminates losses of small amounts of almost-pure target peptides during sample handling.

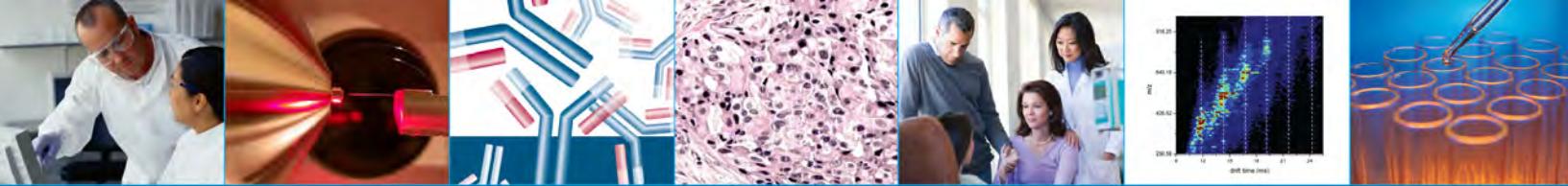
***Optimized and automated peptide capture, wash and elution on Ab-coated magnetic beads for SISCAPA.***

***Invented Accurate Inclusion Mass Screening for qualifying 100's of protein candidates/week in plasma prior to committing to the resource intensive steps of establishing a quantitative assay.***

***Developed a computational algorithm to prioritize lists of cancer-associated proteins for assay development.*** The algorithm integrates multiple “omics” datasets, weights each appropriately, and outputs a rank-ordered list of biomarker candidates.

***Developed a random forest-based algorithm with high predictive accuracy for selecting the best peptides from candidate proteins to use for MRM and SISCAPA-MRM assays.***

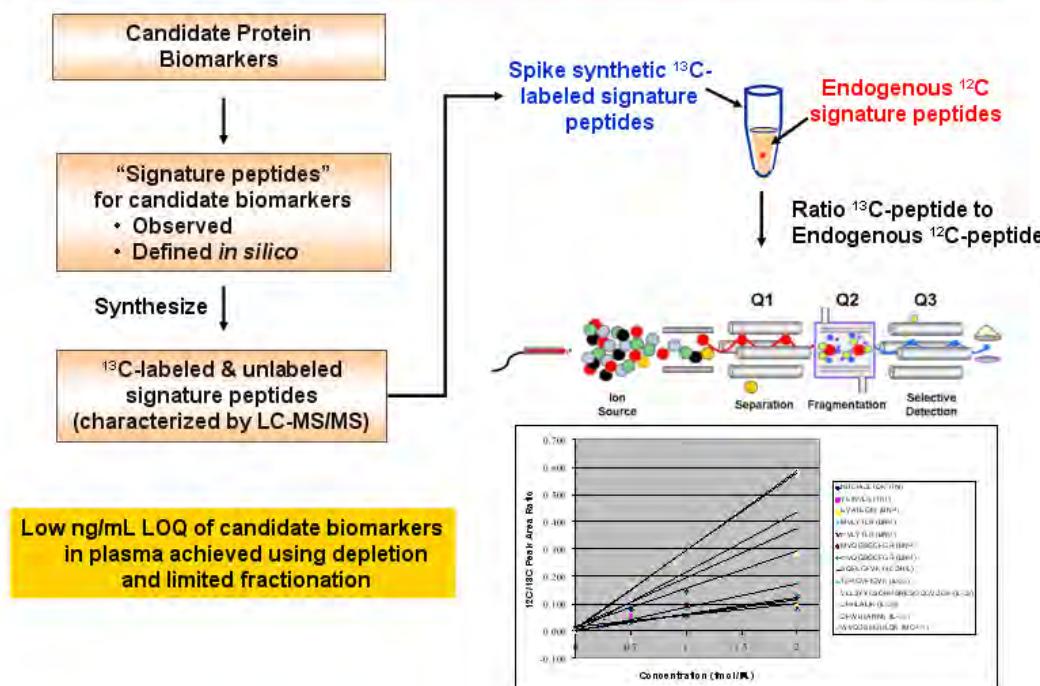
***Communicated problems and solutions to the biomarker research community through 12 published papers (+8 papers submitted), and 74 invited presentations by four program leads during the first two years of CPTAC.*** Communications have included applied methods base to four externally funded biomarker assay projects and four significant collaborations external to CPTAC.

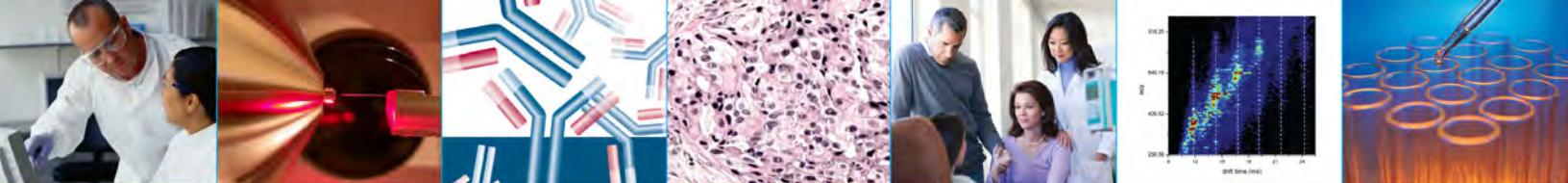


**Figure 25. Verification of candidate protein biomarkers by multiple reaction monitoring MS using stable-isotope dilution**

The triple quadrupole MS system measures precise relative ratios of analyte signature peptides from the target proteins. Calibration curves (bottom right) allow absolute quantities to be estimated.

## Candidates are verified in plasma by quantitative, targeted MS of signature peptides (MRM-MS)

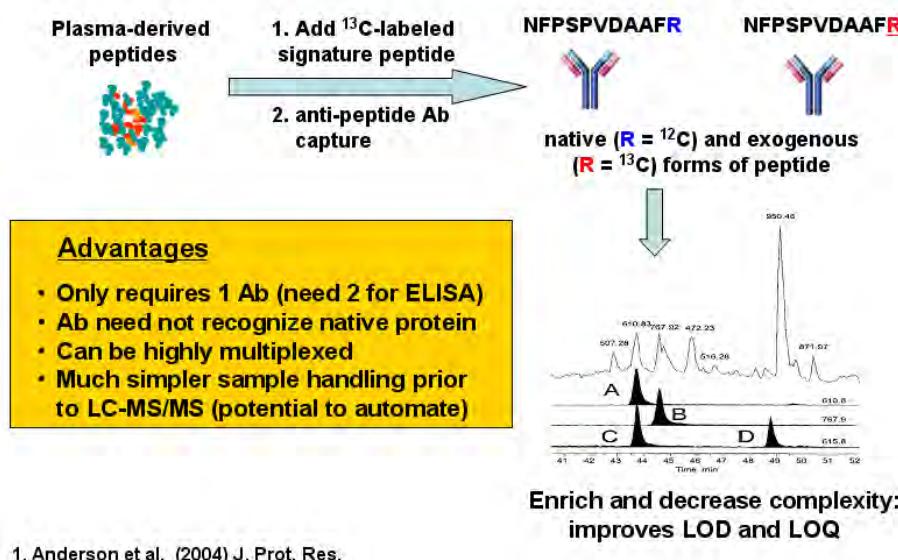


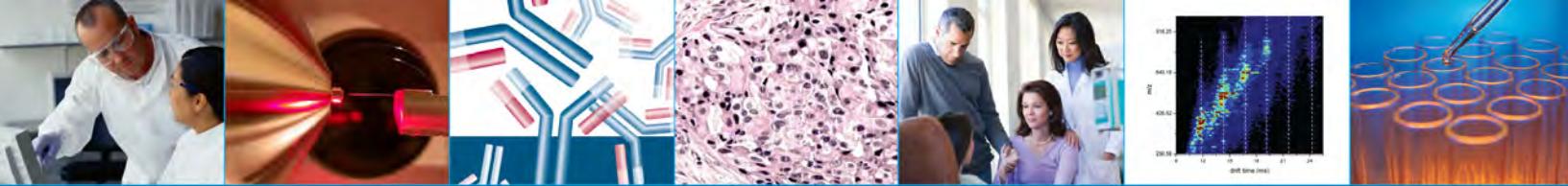


**Figure 26. Schematic illustration of SISCAPA**

MS/MS analysis of peptides bound by antibody to peptide **NFPSPVDAAFR** from a mixture of labeled calibrant peptides and a tryptic digest of whole human plasma. Beneath the total ion chromatogram (top trace) are three traces of ion currents observed in 0.25 amu windows centered at 610.8, 767.9, and 615.8 amu, respectively. Peak A is the **NFPSPVDAAFR** peptide derived from Hemopexin (the assay target) in the digest. Peak B is an incomplete tryptic peptide including **WKNFPSPVDAAFR**. Peak C is labeled **NFPSPVDAAFR**, the isotopically labeled version of the target peptide. Peak D is a peptide derived from plasma ApoA-I lipoprotein, having nearly the same mass as **NFPSPVDAAFR**.

### Targeted MS with Ab-capture of peptides increases sensitivity and assay robustness (SISCAPA<sup>1</sup>)

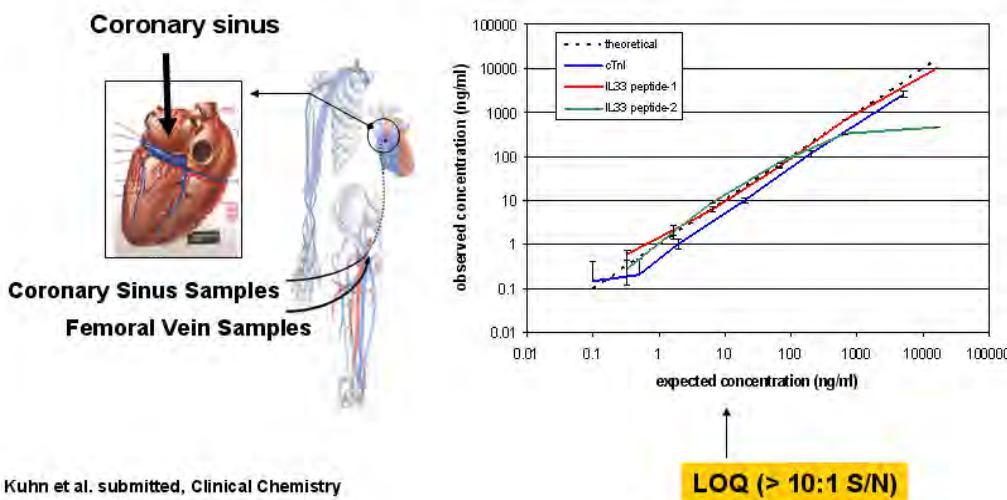




**Figure 27. Multiplexed SISCAPA assay for cTnI and IL-33 has achieved LOQs of ca. 1 ng/mL for both proteins in trypsin digested plasma**

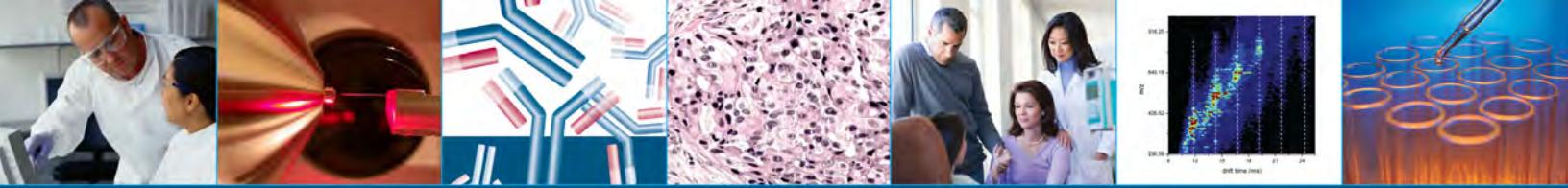
### SISCAPA enables nanogram/mL LOQ of candidates directly from patient plasma

**LOQ of ca. 1 ng/mL directly from 25  $\mu$ L of plasma**



#### Next steps:

The Broad team's long-term goal is to develop technology for quantifying cancer-relevant proteins in biological specimens. During the first two years of the program, we have optimized the performance of SISCAPA-MRM technology via a series of collaborative studies using common reagents (described above). The team's combined knowledge has culminated in a series of SOPs for both MRM and SISCAPA-MRM assays. In the next year, we will continue to test the inter-laboratory performance of MRM with appropriate sample processing and will test the inter-laboratory reproducibility of SISCAPA-MRM assays using a common set of assay reagents. A significant number of critical antibody and peptide standard reagents will become available in the coming year, and we will use these to further develop the SISCAPA assay system and optimize a robust high-throughput implementation suitable for analysis of the CPTAC clinical specimens. In subsequent years, the team will configure assays to cancer-relevant proteins.



## 2.4 Mass Spectrometry Platforms for Post-Translational Modification Characterization

### 2.4.1 Proteomic Characterization of Alternate Splicing and cSNP Protein Isoforms

**Principle Investigator:** Nathan Edwards, Ph.D., Georgetown University

#### **Project Goals and Significance:**

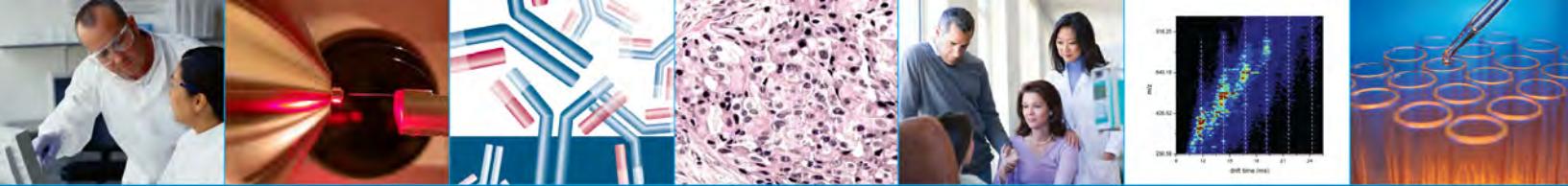
The characterization of alternative splice and variant protein isoforms is a fundamental limitation of current proteomic workflows. To address this issue, Dr. Edwards is developing an infrastructure to enable characterization of alternative splicing and coding isoforms of single nucleotide polymorphisms.

#### **Accomplishments:**

**Peptide Sequence Databases (E).** The peptide sequence databases have made it possible to observe evidence of alternative splicing, cSNP, and novel protein isoforms that otherwise could not be observed in proteomics workflows. Dr. Edwards has described a variety of novel peptides and their implied isoforms that can be observed by the use of this technique. The current peptide sequence database infrastructure, which includes all available sources of protein and mRNA evidence for specific model organisms, removes the implicit bias towards well-understood protein isoforms resulting from the selection, up front, of a protein sequence database for MS/MS searching. The peptide sequence databases are rebuilt every few months and are freely available for download from the PI's website (<http://edwardslab.bmcb.georgetown.edu/>).

**Novel Techniques for Peptide Identification from MS/MS Spectra (E).** The HMMatch and PepArML tools demonstrate that it is possible to *confidently assign many more peptide identifications* than is possible with a single search engine's score or E-value, with no loss of statistical significance. These additional peptide identifications improve protein coverage, the reliability of protein quantitation by spectral counting, and increase the likelihood of observing evidence for peptides that elucidate the distinguishing sites of protein isoforms. PepArML is currently the only open-source alternative to the increasingly popular commercial search engine combiner Scaffold. This approach, called PepArML, has been made freely available as open-source software (<http://peparml.sourceforge.net>).

**Public PepArML Meta-Search Engine (E).** The PepArML Meta-Search engine provides access to large scale MS/MS sequence database searching infrastructure to researchers and labs without the computational resources or personnel to implement a distributed computing strategy in-house. Furthermore, this infrastructure provides a mechanism for search engines not intended or designed to run in a distributed or parallel fashion to be used in a distributed environment, without the need to modify the individual search engines — eliminating the need for the many ad hoc distributed computing solutions embedded in each individual search engine. Lastly, the meta-search engine is designed to be self-contained, platform independent, and require minimal operating system support, making it suitable for installation in small to medium size labs with little



distributed computing expertise. The public service is available from the PI's web-site <http://edwardslab.bmcb.georgetown.edu/pymsio/>).

**Public Peptide Mapping Service (E).** The public peptide mapping service fills a void in the ad hoc mapping of peptide sequences to their source evidence. The blast tools blastp and tblastn are less than ideal because they perform local alignments, dropping amino-acids from the start and end of the query if it improves the alignment statistics. In addition, Tblastn does not provide the nucleotide sequence of the alignment. BLAT, designed for mapping sequences to genomes, requires much longer amino-acid sequences than is typical for tryptic peptides. The public peptide mapping service resolves these issues, and is able to map peptides to source sequences while retaining the underlying nucleotides, and if necessary, project these alignments onto genomic sequence as an annotation track in the UCSC genome browser. This tool facilitates the ability to understand identified peptide sequences in the light of all the available protein sequence and genomic evidence for polymorphism, sequence conservation across species, and translation and splicing.

#### Future Work:

- A plan to converge each of these research efforts into a unified web-based resource for the cancer proteomics community will be developed. This resource will incorporate more novel tools, including spectral matching and the Pevzner Lab's MS Generating Function approach to statistical significance of peptide identifications.
- Ongoing production of the peptide sequence databases from protein or mRNA sequence evidence will be augmented with sequences from *de novo* gene prediction algorithms.
- The Edwards lab will apply their research tools to search public and privately held datasets to look for evidence for known and novel protein isoforms in cancer.

#### Academic/Industry/Research Collaborations:

Fenselau Lab (Department of Chemistry and Biochemistry, University of Maryland)

Liebler Lab (Vanderbilt Medical Center)

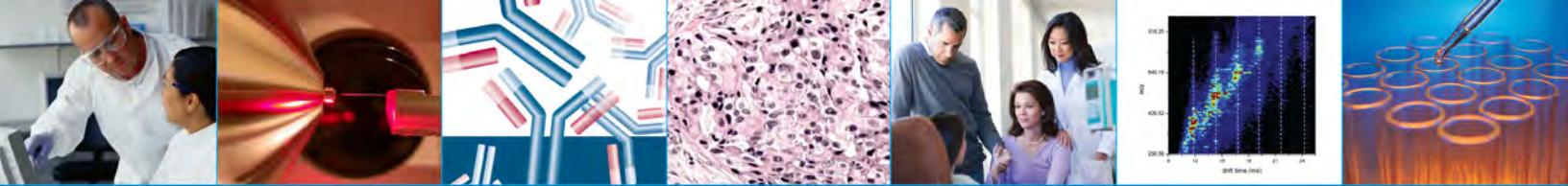
Nesvizhskii Lab (Department of Pathology, University of Michigan)

Goldman Lab (Department of Oncology, Georgetown University Medical Center)

Qian Lab (Beijing Proteome Research Center)

Li Tao (Chinese HUPO)

Pratik Jagtap (Minnesota Supercomputing Institute, University of Minnesota)



## 2.4.2 A New Platform to Screen Serum for Cancer Membrane Proteins

**Investigator:** Daniel B. Martin, M.D., Institute for Systems Biology

**Project Goals and Significance:**

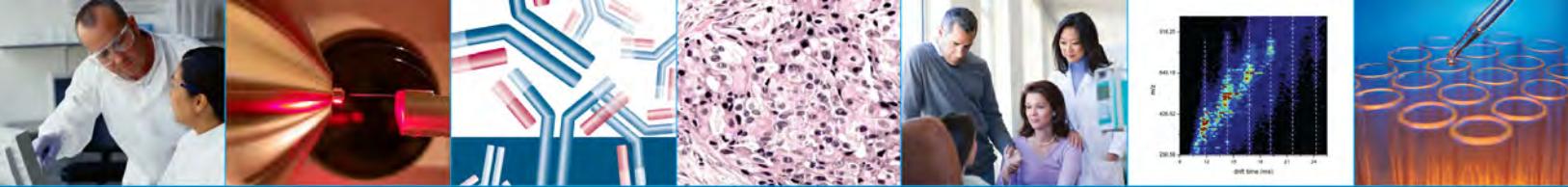
Dr. Martin has optimized a method for the capture of membrane proteins from adherent prostate cancer cells grown in culture. Optimization of the glycocapture conditions includes a variation of the oxidation and covalent biotinylation conditions to minimize cell toxicity. Additionally methods have been developed to exclude glycolipids present on the membrane. Quality control assays have been developed to ensure complete depletion of biotinylated glycopeptides. Using these methods, between 150 to 430 unique glycoproteins have been identified from four prostate cancer cell lines.

**Accomplishments:**

**Created Prostate Cancer Database.** A large database has been created for over 500 glycoproteins and 2000 peptides which were found in prostate cancer cell lines and xenograft tumors. This data serves as the target source for future efforts to identify cancer derived biomarkers in plasma.

**Designed QCAT protein expression system.** A protein expression system was designed to allow the expression and purification of recombinant isotopically heavy QCAT proteins in a yeast auxotrophic system. Using this system, a method was developed to purify recombinant biotinylated proteins to homogeneity using a rapid two step process.

**Developed MaRiMba software package.** The MaRiMba software tool has been developed to automate the creation of explicitly defined MRM transition lists required to program triple quadrupole mass spectrometers in such analyses. MaRiMba creates the MRM transition lists from pre-existing or custom spectral libraries, restricts output to specified proteins or peptides, and filters based on precursor peptide and product ion properties. MaRiMba can also create MRM lists containing corresponding transitions for isotopically heavy peptides, for which the precursor and product ions are adjusted according to user specifications. This open source application is operated through a graphical user interface incorporated into the Trans-Proteomic Pipeline (TPP), and it outputs the final MRM list to a text file for upload to MS instruments.

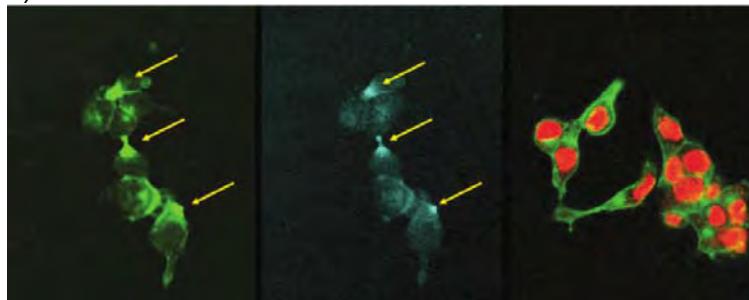


**Figure 28. A) Membrane Co-localization of CD10 and HSP in Prostate Cancer Cells and B)The interface of MaRiMba**

A) *Co-localization of HSP27 and CD10 on C4-2 cells. Immunocytochemistry shows concentration of CD10/HSP27 at certain parts of the cells. Overall, the membrane is stained throughout for CD10 and the cell cytoplasm for HSP27. The 1<sup>st</sup> panel shows CD10, the 2<sup>nd</sup> panel shows HSP27, and the 3rd panel shows both CD10 and HSP27 with nuclear staining (red).*

B) *The interface of MaRiMba. This software allows the creation of MRM-MS analyses based on existing or user-provided spectral libraries.*

A)



B)

1. Specify Input Spectral Library File Show [+]

2. Specify a protein sequence database Show [+]

3. Specify peptide or protein restriction list (optional) Show [+]

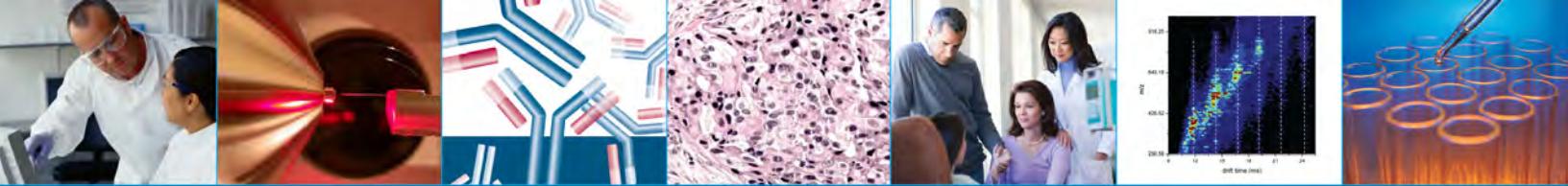
4. Specify Output file  
Output Filename:  
c:/inetpub/wwwroot/ISB/data/dbase/MaRiMba\_18mix\_MRMlist.txt

5. Filtering Options (Show / Hide)

Precursor peptide  
Maximum Number of Transitions per Peptide: 7  
Allowable charge states:  +1  +2  +3  +4  +5  
m/z: Min 300 Max 1200  
pI: Min Max  
Exclude Residues: - - - - - - - -  
Exclude N-terminal Residues  Q  E  
 Exclude all modifications except iodoacetamide Cys C[160]  
 Add transitions for modified (labeled) residues:  
Residue / Mass Difference: - 0.0 Residue / Mass Difference: - 0.0  
Residue / Mass Difference: - 0.0 Residue / Mass Difference: - 0.0  
Residue / Mass Difference: - 0.0 Residue / Mass Difference: - 0.0

Product ion  
Allowable charge states:  +1  +2  +3  +4  +5  
Allowable Ion Types:  b  y  
 Allow Neutral Losses  
 Allow Secondary Small Neutral Losses (e.g. water or ammonia)  
 Allow non monoisotopic peaks  
 Allow mass-shifted ions  
Fragment ion lengths to exclude:  1  2  3  4  5  6  7  8  9  10

6. Generate MRM transition list  
Run MaRiMba



### Next Steps:

- Target prostate cancer xenograft-derived glycopeptides in the plasma of mice using MRM-MS. The information obtained from the membrane proteome of prostate cancer cells may lead to the development of diagnostic tests for human disease. Methods are also being developed that will advance the ability to use the mass spectrometer as a quantitative measurement tool for biological and hopefully clinical samples.

### Academic/Industry/Research Collaborations:

Fred Hutch Cancer Research Center

#### 2.4.3 Quantitative Methods for Spectral and Image Data in Proteomics Research

**Investigator:** Timothy Randolph, Ph.D., Fred Hutchinson Cancer Research Center

#### Project Goals and Significance:

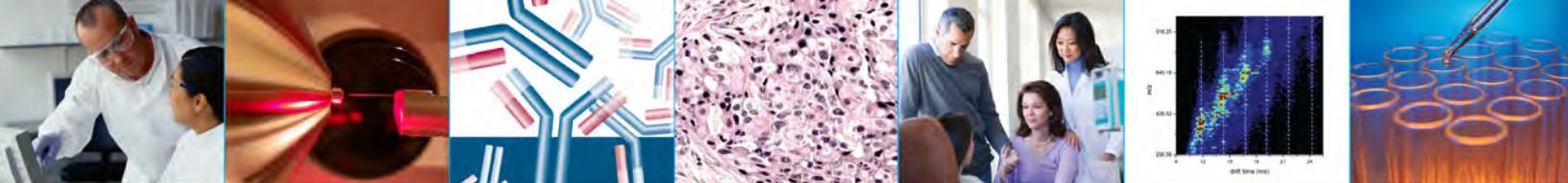
Proteomic-assisted investigations of molecular and genetic function and discoveries of disease-related biomarkers rely on a growing list of technologically advanced assays. Dr. Randolph's efforts are focused on the quantitative and statistical issues inherent in many such platforms. One high-profile platform is LC-MS/MS, but many other platforms—antibody arrays, capillary electrophoresis, high-resolution MALDI MS, immuno-histochemical labeling and vibrational spectroscopies—also serve the goal of identifying proteins relevant to cancer processes and offer different strengths. Major efforts of this work have focused on the analysis of LC-MS/MS, high-resolution MALDI MS and vibrational spectroscopies.

#### Accomplishments:

**Developed an algorithm for detecting differentially expressed proteins between case and control groups.** This algorithm is based on semi-quantitative outputs from label free LC-MS/MS experiments. The method has increased power across a large range of peptide abundance levels.

**Enhanced ability to detect peptide peaks in mass spectra.** Dr. Randolph has developed a method to detect the statistically significant isotopic envelope of peaks from low-abundant peptides in high-resolution TOF spectra.

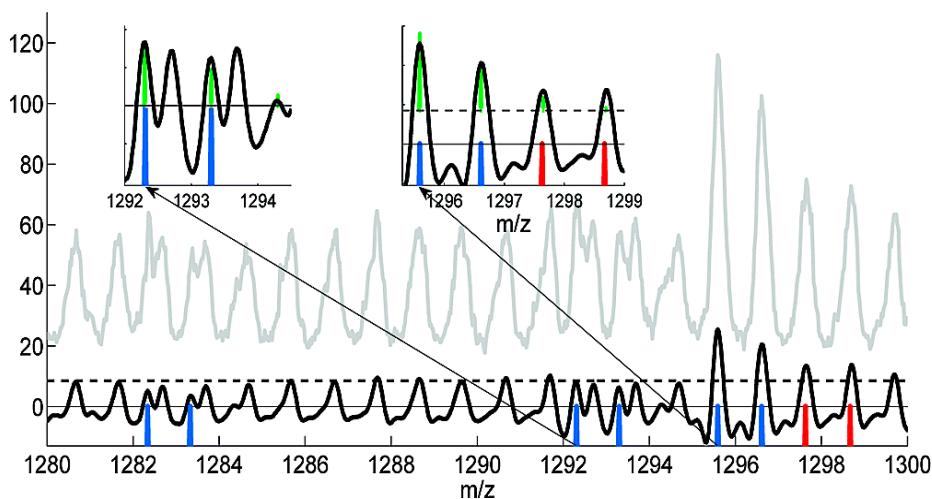
**Developed mathematical formulation of protein-induced signal from high dimensional spectra.** Data from spectroscopies and images are very high dimensional and locally correlated hence require non-standard multivariate statistical methods, such as “functional” regression models. We have developed a rigorous, practical formulation of the mathematics underlying penalized methods for these models. Our research provides a modern understanding of these models which are basic to the analysis of data containing protein-induced signal in the presence of non-protein-related variability.



**Figure 29. Definition of peptide signal in data from high-resolution time-of-flight instruments**

Gray curve, mean spectrum; black curve, mean detail function; dashed line,  $\mu$ l ) mean background-peak intensity; blue tickmarks, location outliers; red tickmarks, intensity outliers. Insets are expanded views, as indicated, with the Poisson density plotted in green.

Image taken from *Journal of Proteome Research*. 2008, 7 (01), pp 276–285

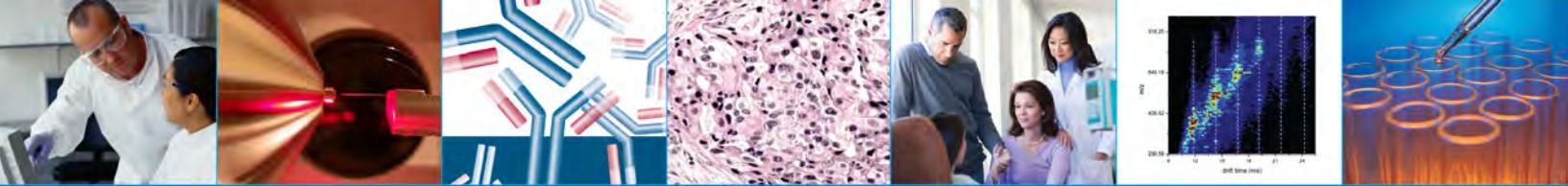


#### Future work:

- Refining, evaluating and comparing the few existing methods for comparative analysis of LC-MS/MS data sets.
- Developing a label-free quantitation algorithm.
- Extending statistical methods for spectral data analysis.

#### Academic/Industry/Research Collaborations:

Dr. Rich Gardner, University of Washington  
Dr. Paul Lampe, Fred Hutchinson Cancer Research Center  
Dr. Larry True, University of Washington  
Dr. Jarek Harezlak, Indiana University School of Medicine



#### 2.4.4 A Proteomics Approach to Ubiquitination

**Investigator:** Junmin Peng, Ph.D., Emory University

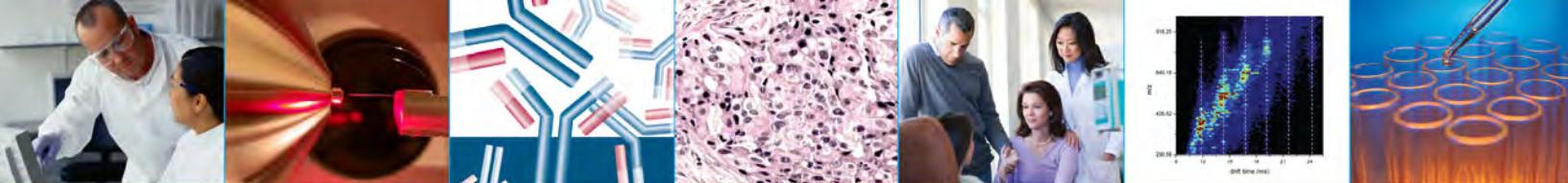
**Project Goals and Significance:**

Dr. Peng seeks to develop a proteomics approach to isolate and quantify ubiquitinated proteome from tumor tissues based on high resolution MS. Ubiquitin modification regulates a variety of cellular events in eukaryotic cells by covalently attaching ubiquitin to target substrates. The dysregulation of ubiquitin pathways has been involved in the pathogenesis of many types of cancers. However, there is no reliable method to globally analyze ubiquitinated proteins in cancer samples. Recently, progress has been made in developing novel MS-based technologies to analyze ubiquitinated proteome in yeast. Dr. Peng hopes to develop a generic method to capture and quantify ubiquitinated proteome from mammalian cells.

The approach will utilize affinity chromatography composed of ubiquitin-binding domains. An array of fusion proteins with various ubiquitin-binding domains will be examined to optimize the binding affinity in order to improve the method with respect to the yield and purity of ubiquitin substrates. The isolated ubiquitinated proteins will be further analyzed by mass spectrometry for the determination of protein identity and ubiquitination sites. Quantitative mass spectrometry strategies will be implemented to investigate the dynamics of ubiquitinated proteome. Once established, the method will be highly useful for profiling ubiquitinated proteome in mammalian samples, including clinical tumor tissues.

**Accomplishments:**

**A novel method for removal of false positives from the ubiquitinated proteome.** Two approaches to isolate ubiquitinated proteins from cell lysate have been developed: (i) by affinity purification based on tagged ubiquitin, and (ii) by affinity isolation through ubiquitin-binding domains. Regardless of the method, one challenge for analyzing ubiquitinated proteome is how to differentiate genuine ubiquitin-modified proteins from co-purified contaminants (i.e. false positives). Therefore, a strategy was developed to validate ubiquitinated proteome based on 1D SDS gel followed by liquid chromatography and tandem mass spectrometry (GeLC/MS/MS). This strategy allows the reconstruction of a “Virtual Western Blotting” image for every protein identified by MS, thus detects large mobility shift of ubiquitinated species during gel electrophoresis. Using this strategy, it was found that only approximately 30% of proteins were truly modified by ubiquitin in affinity-enriched samples. The strategy provides a simple, effective method for quality control in analyzing ubiquitinated proteome.

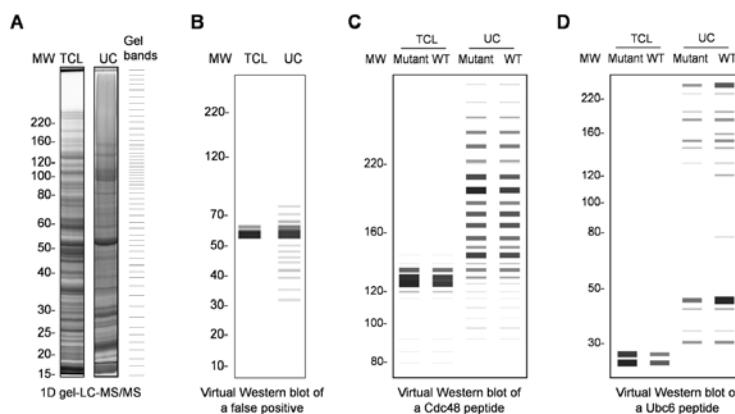


**Figure 30. Qualitative and quantitative analysis of ubiquitinated proteome by mass spectrometry**

(A) Comparison of the total cell lysate (TCL) and the enriched Ub-conjugates (UC) by SDS-PAGE. Both samples were resolved on a gel, excised into ~50 gel bands, and analyzed by LC/MS/MS.

(B) A representative false positive protein revealed by reconstructed virtual Western blotting. The protein abundance was represented by the darkness and thickness of the bands; and the molecular weight (MW) information was extracted from the 1D SDS gel.

(C-D) Virtual Western blots for Cdc48 and Ubc6, quantified in wild type (WT) and mutant yeast strains.



**Discovered an important regulation mechanism for protein ubiquitination.** In addition, the proteomics analyses of ubiquitinated proteome revealed a surprising diversity of polyubiquitin structures. A bottom-up method was first developed to quantify the polyubiquitin linkages using synthetic peptides as internal standards, and then a middle-down mass spectrometry method was further developed to dissect the polyubiquitin structure. This approach was used to profile total cell lysate and ubiquitinated proteome in a single experiment, and an important regulation mechanism for ubiquitination was discovered. Importantly, these proteomics methods have been utilized to probe the function of protein ubiquitination, and it has been discovered that all non-K48 polyubiquitin chains are critical for ubiquitin-proteasome system function.

#### Next Steps:

- Continue to develop the proteomics platform for analyzing ubiquitinated proteomes, and apply the platform to mammalian cells and clinical samples.

#### Academic/Industry/Research Collaborations:

Harvard University

Georgia Institute of Technology

Massachusetts Institute of Technology

University of Michigan

University of Pennsylvania

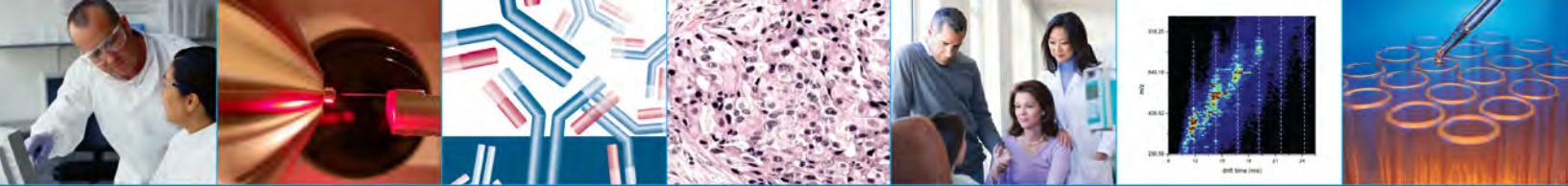
Tufts University

Auburn University

Yale University

McGill University

Emory University



#### 2.4.5 A Cell-Free System for High-Yield Phosphoprotein Synthesis

**Company:** Rana Biosciences  
**SBIR**

**RanaBio**

**Principal Investigator:** Guo-An Wang, Ph.D.

**Proposal: A cell-free system for high yield phosphoprotein synthesis**

Protein phosphorylation is a central mechanism of cellular regulation. Aberrant phosphorylation activities can cause human cancers. Highly specific capturing reagents are needed to detect and monitor these activities with the aim of detecting cancer early. To generate, characterize and validate these capturing reagents, sufficient amounts of functional phosphoproteins are needed. The long-term goal of this project is to establish a cell-free protein synthesis system that can be used to produce large quantities of site-specifically phosphorylated proteins. The system will feature low cost, high reproducibility, high quality and high yield. The key advancement of the new method is made possible by incorporating phosphoamino acids through site-directed nonsense suppression. Specific reagents will be developed to activate phosphoamino acids efficiently and continuously in the translation reaction. Translation termination at the selected nonsense codon will be manipulated to achieve high suppression efficiency. This cell-free protein synthesis system will comprise of both prokaryotic and eukaryotic translation systems. In phase I, the objective is to develop an *E.coli* cell-free system that is able to produce proteins containing phosphotyrosines in a yield up to 0.5 mg per 1 ml of translation reaction.

**Accomplishments:**

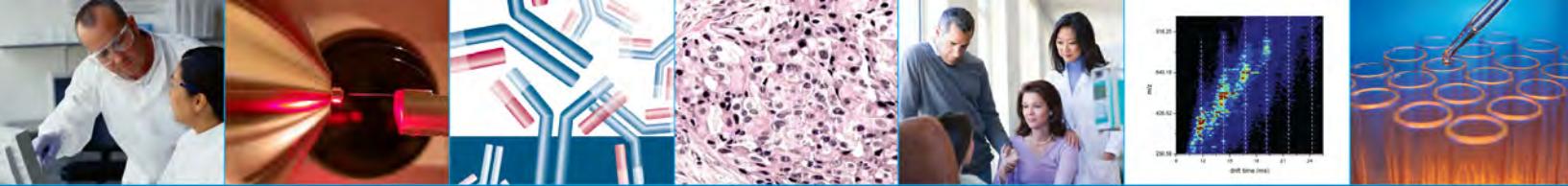
During phase I period of this project, Rana Biosciences has set up conditions for both *in vivo* and *in vitro* selection of synthetases. They have developed an effective screening strategy and achieved 96% suppression efficiency by use of release factor (RF1) antibody. Rana Biosciences has identified active synthetases for phosphotyrosine although additional tests are needed.

#### 2.4.6 CPTAC Post-Translational Modification Workflows

**Principle Investigators:** CPTAC round robin study

**Project Goals and Significance:**

In complex biological systems, any given cellular or extracellular protein is likely to exist as an ‘ensemble’ of proteins made up of a related set of isoforms. In particular, any given protein may have several splice variants, undergo extensive enzymatic PTMs, or be oxidatively damaged. These ensembles of modified proteins represent an important opportunity for biomarker discovery as cancer processes are known to involve significant changes in protein splicing, glycosylation, acetylation, phosphorylation, proteolysis or



redox status. It is therefore the purpose of these CPTAC round robin studies to develop bio-analytical strategies where these post-translational or post-transcriptional modified proteins can be specifically targeted for plasma biomarker discovery using MS-based platforms. The objectives are to:

- Develop, optimize and implement workflows and mass spectrometry platforms for analysis of PTM proteins in plasma
- Develop appropriate protein and/or peptide standards for the evaluation of PTM-based workflows and mass spectrometric analyses

#### Accomplishments:

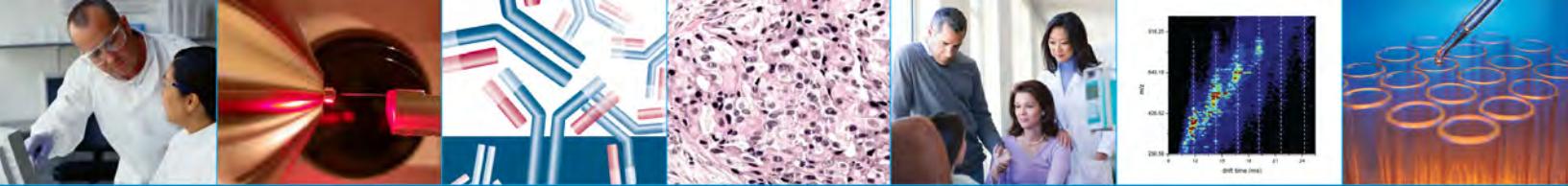
- Developed a comprehensive plan for glycoproteomics that included both protein and peptide-based enrichment strategies. The initial part of this plan was developed among the Purdue, UCSF, and Buck research groups in the first half of 2008. Experiments included using 5 different agarose-bound lectins (AAL, LEL, SNA, UEA-1 and ConA) to enrich in specific sets of plasma glycoproteins. These proteins were analyzed before and after lectin enrichment with both the captured and flow-through fractions subjected to 1D gel separation, followed by MS to assess the complexity and efficiency of this affinity enrichment process. Both MALDI-MS (4800 TOF/TOF and vMALDI-LTQ) and ESI-MS (QSTAR) platforms were used to analyze the gel separated protein fractions, yielding 20 to 50 glycoproteins for each lectin when MALDI was used, or over 40 proteins for 12 selected bands when analyzed by HPLC-ESI-MS.

Experiments were implemented to optimize and standardize this approach. The second part of the proposed Glycoproteomics study will have each participating CPTAC laboratory essentially start with depleted plasma and work up the sample at their site. Each site will be given the necessary reagents to carry out these experiments.

- Optimizing a comprehensive workflow strategy for targeting phosphoproteins and oxidative proteomics.

#### Next Steps:

- Launch the optimized glycoproteomics workflow in a CPTAC wide study.
- Consider the addition of iTRAQ labeling to the PTM workflows, especially as this will allow 8-plex labeling.
- Work with the unbiased discovery and verification workgroups to design and launch new PTM-experiments. In particular, a glycopeptide-based MRM experiments that could target deglycosylated peptides would be important to move forward
- Continue to develop and examine workflows for all PTM targets, but with a new focus of phosphopeptide-based plasma proteomics.
- Continue to work with the ANL to evaluate a plan for developing <sup>15</sup>N-labeled phosphoproteins via treatment with specific kinases.



- Identify means to acquire various labeled PTM-peptides as standards
- Examine the possibility for *de novo* synthesis of target proteins containing various PTMs.
- Examine the utility of off-line MALDI-MS platforms for targeted PTM-peptide analysis (glycopeptides, phosphopeptides and oxidized peptides).
- Work with the bioinformaticians (*i.e.*, Ron Beavis at UBC) to develop a fully interactive, annotated protein resource that supplies key information regarding the PTM status of plasma proteins.

## 2.5 Immunoaffinity Platforms and Separation Technologies

### 2.5.1 Analytical Proteomics

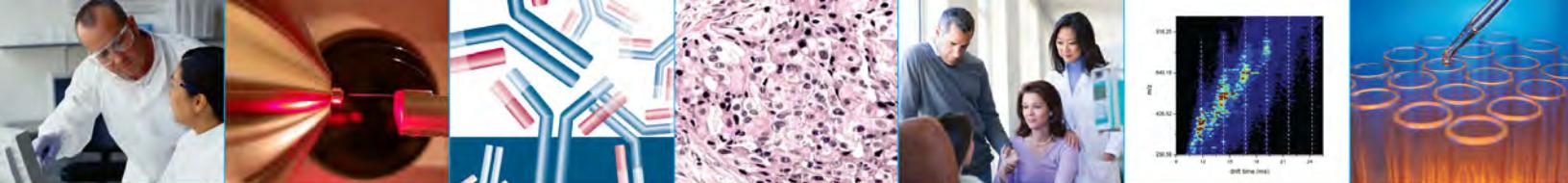
**Principal Investigator:** Fred E. Regnier, Ph.D., Purdue University

**Research Interests:**

- High-throughput immunoaffinity and separation technologies and MS instrumentation
- Microarray technologies employing interferometric analysis

**Project Goals and Significance:**

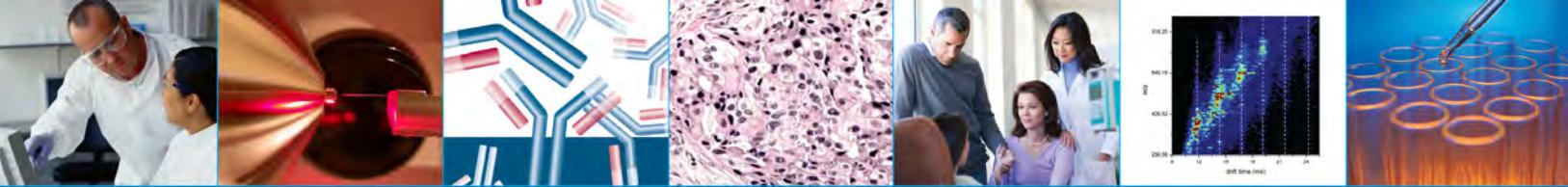
MS-based proteomics approaches have provided insight into biomarkers of cancer and other diseases with femtomole sensitivity and high analytical precision but the results have proven difficult to reproduce owing to the complexity of human biofluids and to the protocols employed in these approaches. This CPTAC team develops robust protocols and standards for MS proteomics employing both ESI and MALDI platforms. Ion mobility MS provides several orders of magnitude increased dynamic range and protocols for detection of cancer biomarker candidates will be further explored with this emerging technology platform. The consortium will employ high specificity immunologic reagents to develop platforms for precise detection and quantification of biomarkers of relevance for breast and prostate cancer. Immunoaffinity selection of specific biomarker proteins will be achieved as a sample fractionation step using nano-scale immunoaffinity columns prior to MS detection and quantification of specific proteins. An emerging technology incorporating specific antibodies on a microfabricated 'bioCD' read by spinning disc interferometry enables label-free evaluation of hundreds of analytes from hundreds of samples in minutes. The team will focus on candidate biomarkers relevant for breast and prostate cancer from the NFkB and STATS signaling pathways. Three for-profit and one not-for-profit corporate partners will join several academic groups with deep expertise and extensive resources in proteomics technologies to most efficiently evaluate and roll out robust protocols and standards for MS- and affinity proteomics approaches. Prostate cancer samples will be made available from the NCI-sponsored ECOG trial; breast cancer patient and control samples will be collected specifically for this consortium by the Hoosier Oncology Group from throughout the state and region. Standardized sample collection organized from the Indiana University School of Medicine, will ensure a diverse cross section of patient demographic groups. Three additional corporate partners will provide key technologies and consultation to improve team efficacy.



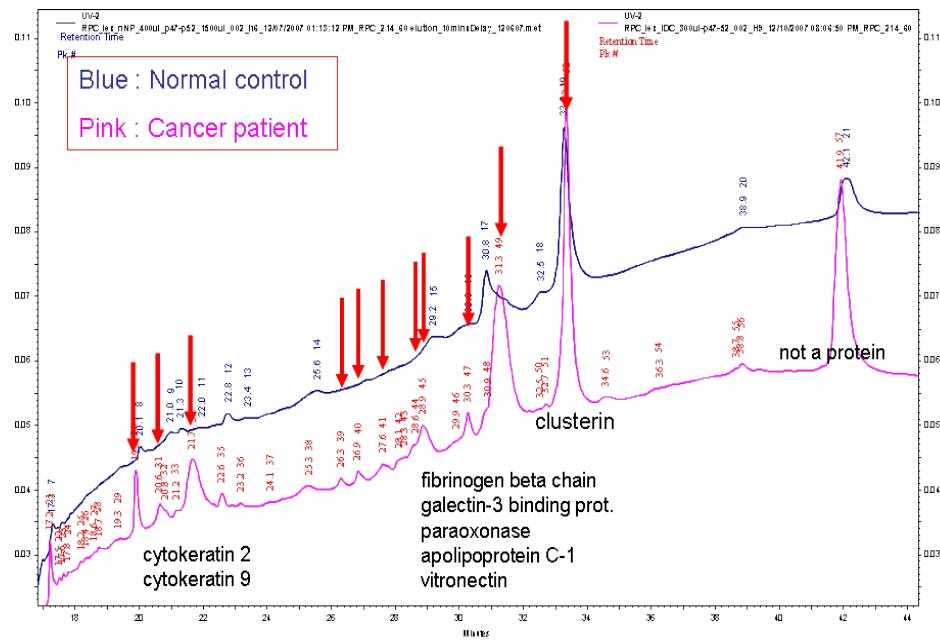
### Accomplishments:

**The team investigated the premise that “the performance of platforms for clinical proteomics can be vastly improved through the selection of more appropriate, or improved separation and enrichment technology ahead of detection systems, irrespectively of the particular type of detection system being used”.** Two types of new, or improved separation and/or enrichment systems were examined in this work:

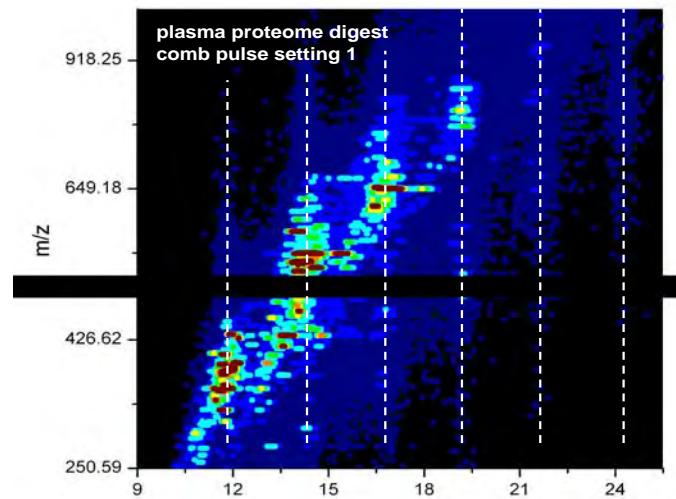
- A) Affinity selectors that target broad features of intact cancer marker proteins;
  - B) Ion mobility separators.
- **Affinity Selectors for Modified Proteins:** Results from A) indicated that the speed, specificity, and enrichment afforded by antibodies and lectins is an extremely powerful asset that greatly increases the performance of any platform using affinity selectors to pretreatment samples before final detection (Figure 31). This asset is so important that affinity selection will probably be a component of all future clinical proteomics platforms. This work has been published in Analytical Chemistry (Cho et al, 2008).
  - **Enhancing Mass Spectrometry Throughput of Polypeptides:** Results from B) concluded that IMS-IMS-MS i) provides up to 60 times faster (3 min) fractionation and analysis of polypeptides from plasma proteins than conventional reversed phase chromatography-mass spectrometry ii) with only a 20 to 25% reduction in the number of proteins identified iii) while achieving slightly higher sensitivity (Figure 32). A further conclusion is that in a clinical setting the performance of an IMS-IMS-MS platform will be vastly superior to LC-MS systems.

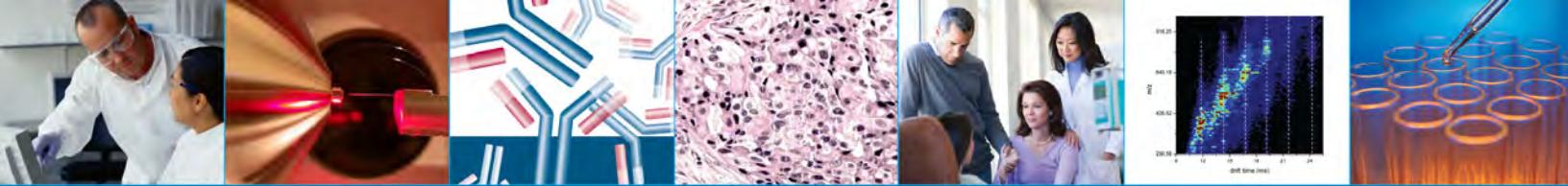


**Figure 31. RPC chromatography of Lewis-x containing glycoproteins enriched by using immobilized IgM immunoaffinity chromatography (IAC) columns in plasma**



**Figure 32. Two-dimensional (drift time, m/z) dot plot of combing pulse**  
*Feature intensities are represented by a color map. Mobility selections (comb teeth) are represented by dashed lines.*





**Investigated standardization of separation/enrichment platforms.** Multiplexed affinity chromatographic (MAC) systems including immunoaffinity chromatography (IAC) and RPC platforms require very different mobile phases, elution conditions, kinds of columns, and types of gradients. The team has study parameters that will help standardize the platforms.

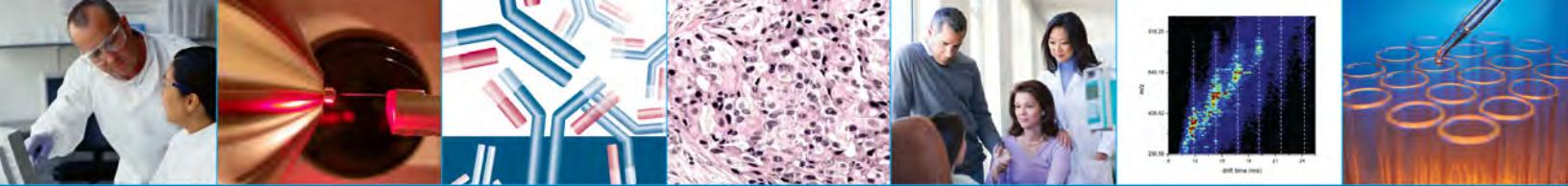
**Assessed reproducibility of peptide identification in shotgun proteomics at two levels: experimental and computational.** The NCI-20 sample was used to estimate inter- and intra-laboratory experimental reproducibility.

**Devised a clinical proteomics data model for managing metadata of mass spectrometry experiments.** In order to effectively describe and prototype the experimental metadata for tandem MS clinical proteomics, we designed and implemented Entity-Relationship data modeling to capture and manage experimental designs, experimental protocols, data storage parameters, search software, and annotation of proteomics result descriptions. This data model has been implemented as a customized application of the metadata framework in the CPAS.

**Designed online Healthy Human Individual's Integrated Plasma Proteome (HIP2) Database.** One of teams' efforts toward the standardizing the proteomic data analysis among multiple laboratories was to create a peptide annotation database for the proteomics community, where biomedical researchers can search for plasma proteins collected from different MS platforms together with experimental protocols and protein identification software for healthy individuals. This web-based database will be useful to biomedical researchers involved in biomarker discovery research. The primary goal of the HIP2 database is to support future clinical proteomics research, especially the discovery of biomarkers through plasma proteomics profiling.

**Developed a hierarchical statistical model to assess the confidence of peptide and protein identifications made by tandem mass spectrometry.** The team developed a statistical framework to improve the identification accuracy and reliability of false positive estimate for peptide/protein identifications in shotgun proteomics. The model in its essence is a hierarchical statistical model (HSM) in the empirical Bayes framework that describes the experimental process that generates the scores. This model offers a general framework that can be applied to any continuous scores. Overall, the HSM demonstrates promising results.

**Devised a Bayesian approach to protein inference problem in shotgun proteomics.** The problem of determining which of the proteins are indeed present in a sample, known as the *protein inference problem*, represents a major challenge in shotgun proteomics. The team address this problem by proposing a novel Bayesian approach that takes as input a set of identified peptides from any peptide search engine, and attempts to find a most probable set of proteins from which those identified peptides originated. The model was tested and the results suggested that Bayesian inference model outperforms ProteinProphet by reporting more true positive proteins, less false positive proteins, and less false negative proteins.



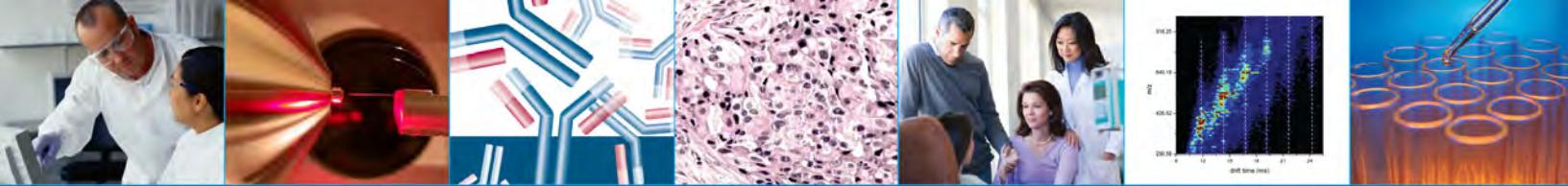
***Progress made toward development and implementation of uniform algorithms for sharing bioinformatics and proteomic data and analytical/data mining tools across the scientific community:***

- Developed a web-based Proteome Discovery Pipeline (PDP) for sharing the proteomic data analysis tools among multiple laboratories. This web-based analysis platform provides for complete proteomics data analysis without requirement for specialized hardware or input from bioinformatics specialists.. Key advantages of PDP include the capability of processing experimental data generated by either label-free and/or stable isotope labeling proteomics; the ability to accept multiple experimental data formats; and the flexibility of incorporating multiple algorithms for each data analysis step using web services. This extensible system has been successfully utilized for several protein biomarker discovery projects.
- Assessed an approach for prediction of peptide fragmentation spectra. Using spectral library for peptide identification, the team attempted to generalize this approach by accurately predicting the theoretical spectrum of a peptide from its sequence, not only for the occurrences of the fragment ions but also their relative intensities. Combining a spectral library approach with accurate fragmentation prediction has the potential to provide complementary tools of the highest accuracy for peptides that have either been previously observed (spectral library) or not (spectral prediction).

***Progress is being made toward developing well defined and comprehensively characterized sets of standard/reference materials and reagents to serve as resources for the research community.*** Generation of standards are being concentrated toward Carbonylation, Glycation, Phosphorylation, Glycosylation.

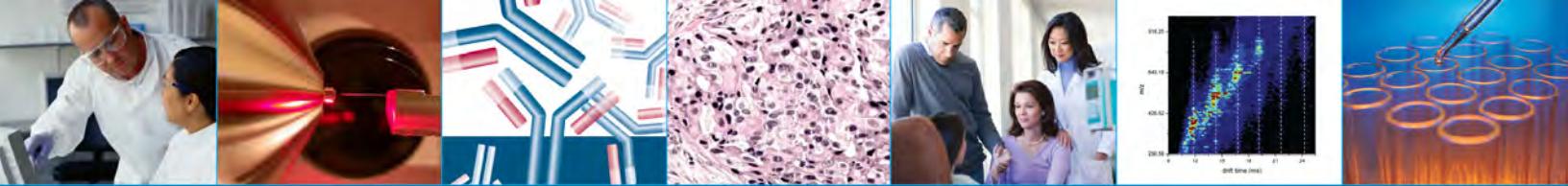
***Evaluated proteomic platforms for their ability to analyze cancer relevant proteomic changes in human clinical specimens:***

- IMS-IMS-MS performance was further evaluated in the analysis of plasma samples. One set of experiments focused on determining the reproducibility of peak intensities as needed in quantification. The focal point of the study was to determine the degree of variability in plasma proteomics analyses. The team concluded from the studies that the IMS-IMS-MS system is both robust and high-throughput in plasma proteomics applications.
- A MAC analysis protocol was outlined in which IgM antibodies (noted previously) were used to select glycoproteins bearing either Lewis x ( $\text{Le}^x$ ) glycan or sialyl-Lewis x ( $\text{sLe}^x$ ) glycan in plasma samples from control and breast cancer patients. Results showed a marked difference in the concentration of a small number of proteins carrying this antigen between normal subjects and cancer

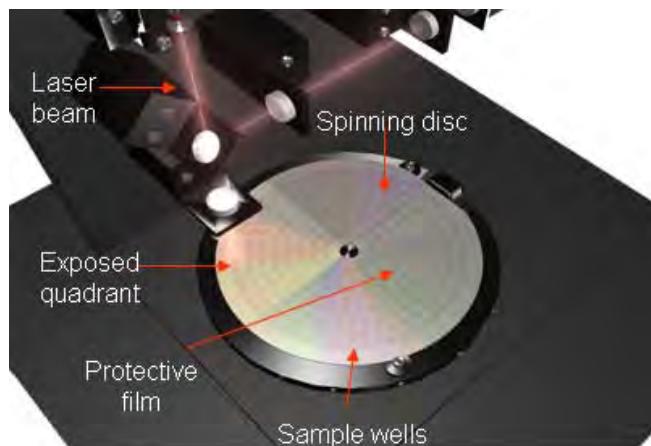


patients. A large scale study is being initiated using the multiplexed affinity chromatography platform in which at least a hundred each of breast cancer patient and controls samples will be examined to proceed on toward validation of these marker candidates.

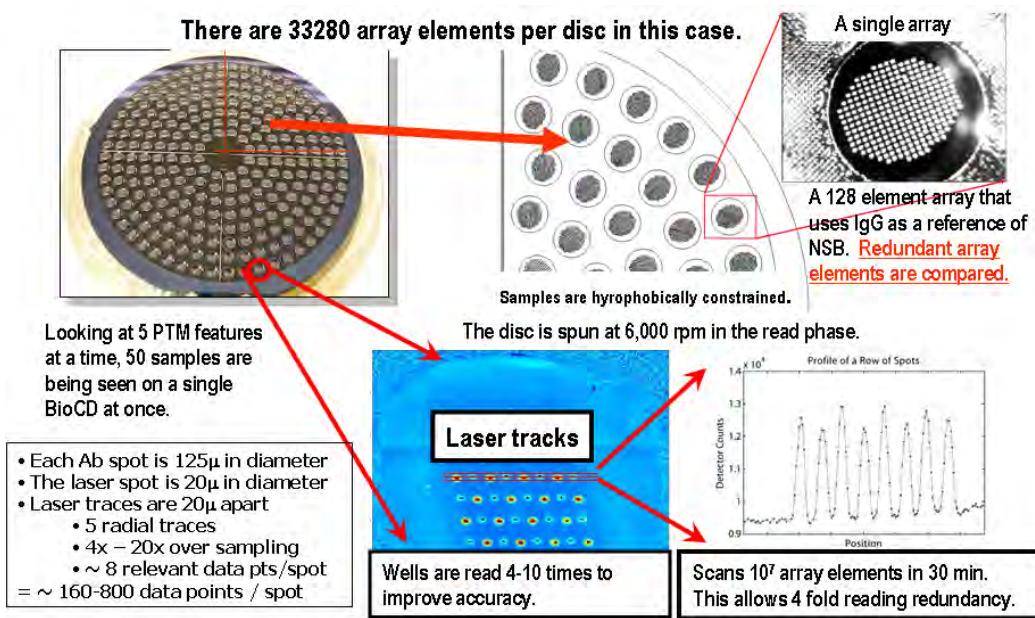
- The team applied a label-free LC/MS-based protein quantification technology to analyze plasma samples from 40 healthy women and 40 breast cancer (stage I & II) women. Using two-sample t-statistics and permutation procedure, the team identified 254 statistically significant differentially expressed proteins, among which 208 are over-expressed and 46 are under-expressed in breast cancer samples. The team validated this result against previously published proteomic results of human breast cancer cell lines and signaling pathways to derive 26 candidate protein biomarkers in a panel.
- Applied animal model of breast cancer to discover potential new biomarkers (collaboration). Candidate metastasis biomarkers were examined in rat lymph with the goal of identifying signature peptides in lymph from rats with metastatic mammary tumors and evaluating the predictive value of these peptides in lymph and serum, including lymph from a set of rats with metastatic mammary tumors before and after surgery and rats with non-metastatic mammary tumors.
- Developing high-throughput Immunological Assays. Immunological assays have been a cornerstone of protein quantification. They are still the gold standard in terms of sensitivity, simplicity, and cost per analysis in clinical chemistry. The Purdue group is in the early stages of evaluating immunological array technology for cancer biomarker quantification. The system being examined is manufactured by Quadraspec Inc. of West Lafayette, IN. At present the Quadraspec high-throughput immunological assay system is being validated for use in translational proteomics. It is anticipated that validation of cancer markers in translational medicine could require the analysis of 10,000 or more patient samples. Because the Quadrapec system is capable of analyzing 500 to 1,000 samples an hour at \$1 to \$5 per sample, it could be of value in validating the diagnostic and/or prognostic efficacy of a multiple protein molecular signature of cancer. A second attractive feature of this technology is that a slightly different, less expensive version of the same instrument can be used to carry out clinical assays with single patients.



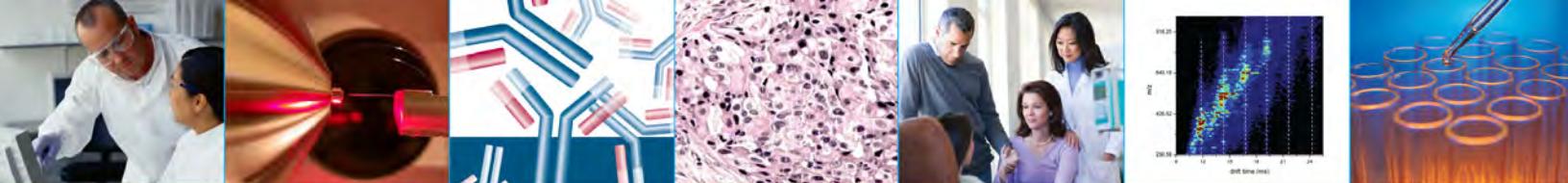
**Figure 33. Immunological array and optical reader**



**Figure 34. Expanded view of antibody arrays on a high-throughput immunological assay disc**



The team significantly contributed to consortium-wide endeavors and has been particularly noteworthy for the progress of the inter-laboratory studies carried out in multiple Working Groups particularly PTM WG. The Purdue team designed the



experiments, helped develop the SOPs and provided statistical guidance for PTM workflows that are currently being tested across multiple CPTAC centers.

#### Next steps:

- Continue marker quantification studies with breast cancer patient plasma samples using the small set of potential markers we have currently identified with the objectives outlined below:
  - Collect and examine samples from at least 200 patients.
  - Determine the degree to which coupling affinity selection strategies to shot gun proteomics, ion-mobility separator MS based proteomics, and high-throughput array platforms improves their performance in quantifying breast cancer markers.
  - Quantify the degree to which cancer marker proteins are found and vary in both normal and cancer subjects.
  - Complete an informatics package that goes from patient sample collection to archiving all relevant data to extensive data analysis.
- Expand team's study of prostate cancer patients in parallel with the ongoing study of breast cancer patients, but with the slightly different objectives listed below:
  - Determine the extent to which platforms and protocols found to be efficacious in breast cancer proteomics apply to prostate cancer.
  - Begin the identification of potential biomarkers in prostate cancer patients.
  - Determine the degree to which affinity selectors described in the literature as being of value in recognizing prostate cancer indeed work, as was the case with team's studies in breast cancer.
- Examine a small set of colorectal, ovarian, and lung cancer patient samples along with arthritis patient samples obtained from commercial sources using the breast and prostate sample protocols to determine the degree to which Purdue team can differentiate breast and prostate cancer from other types of cancer and inflammatory diseases.

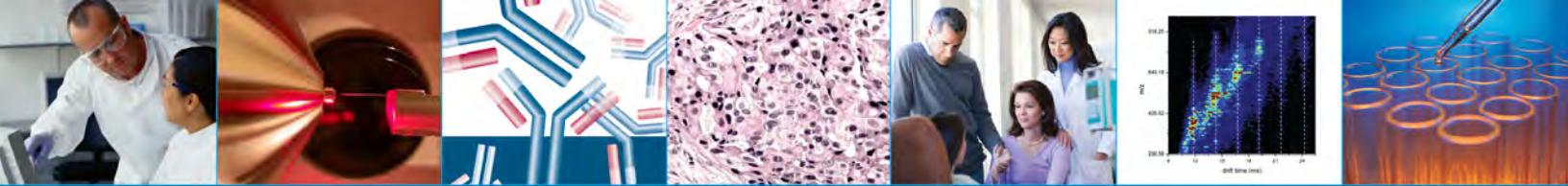
#### 2.5.2 Automated Multiplexed Immunoassays for Rapid Quantification of Low Abundance Cancer-related Proteins

**Company: Rules-Based Medicine, Inc. Austin, Texas  
SBIR, Fast Track**



#### Proposal:

The company proposed to develop a quantitative, automated, 5-plex immunoassay for the rapid detection of low abundance cancer-related proteins in phase I. In phase II, the



Company proposes to validate the screening platform developed in phase I by screening an additional 45 targets. Therefore, a 50-plex immunoassay for low abundance cancer-related proteins will be developed.

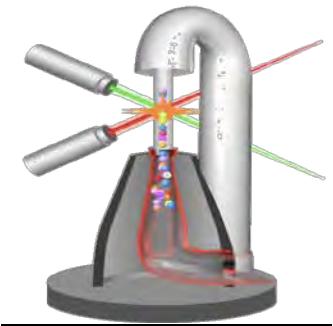
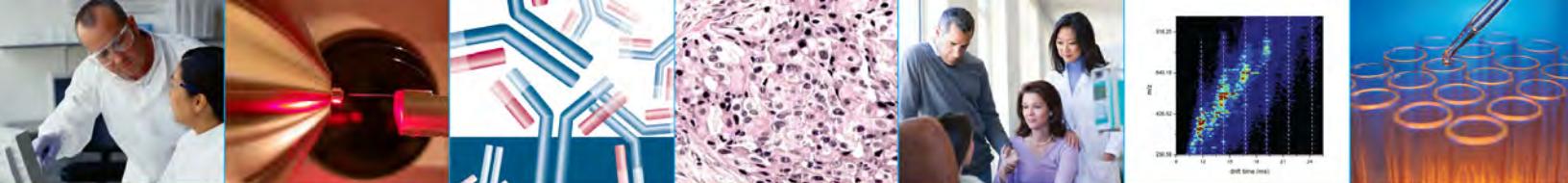
### **Set Milestones:**

#### **Phase I**

- Place reagent orders
- Demonstrate feasibility
  - Develop custom software application for the automated assays
  - Automate multiplex fabrication
  - Automate immunoassay development
  - Automate optimization of each individual assay and characterize assay performance
  - Conduct usability testing
  - Automate immunoassay multiplex development
  - Perform assay validation and determine feasibility
- Establish prototype revisions/additions to be implemented in Phase II
- Present the finding to an NCI Evaluation Panel
- Submit Phase I final report

#### **Phase II**

- Place instrument and reagent order(s)
- Setup, calibrate and validate new instrumentation
- Implement prototype revisions/additions established in Phase I
- Automate cross-reactivity validation
- Implement barcode reading system
- Optimize automated screening platform and processes
  - Reaction plate layout
  - Volumes
  - Worktable layout (speed)
- Couple automation steps developed in Phase I
- Validate reproducibility of screening platform (Development of 50-plex)
- Submit Phase II Interim Report
- Perform assay validation
- 50-plex automated cross-reactivity validation
- 50-plex assay for cancer-related low abundance proteins
- Present finding to an NCI Evaluation Panel
- Document final protocols and methods
- Submit Phase II Final Report
- Submit Annual Commercialization Update



### Accomplishments:

Phase I, combined with historical data on existing Multi-Analyte Profile (MAP) technology analytes has demonstrated the ability to use a sample volume of <50  $\mu$ L to achieve a 5-plex with limit of detection < pg/ml, CVs <10% and a broad dynamic range. These specifications are attainable using RBM's automated sample analysis MAP service as well as the automated screening platform developed in phase I for rapid selection of optimal ligand-binding pairs.

### 2.5.3 Multiplex Mass Spectrometric Immunoassays

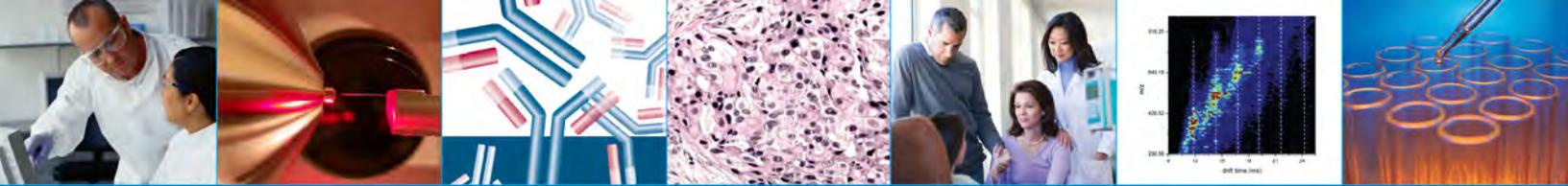
**Company: Intrinsic Bioprobe Inc.**  
SBIR, Fast Track



**Principal Investigator:** Dobrin Nedelkov, Ph.D.

#### Proposal: Multiplex Mass Spectrometric Immunoassays

The objective of this contact in phase I research is to develop and validate multiplex mass spectrometric immunoassays (MSIA) for detection and quantification of cancer-related proteins with intrinsically low bodily fluids concentrations. In the first specific aim, the company will develop, optimize, test, and validate a MSIA assay for cancer-related protein whose intrinsic concentration in a bodily fluid (i.e., plasma) is in the 1 to 10 ng/mL range. A quantitative MSIA assay utilizing a standard curve approach will be developed, and the limits of detection and quantifications will be determined. The assay will be tested with at least a dozen different human plasma samples, in triplicate, and the CVs of the assay will be determined. Usability testing of the assay will be performed with three representative users, and their feedback will be used to make modifications and improvements. The results obtained with the MSIA assay will be compared to a commercially available ELISA assay. In the second specific aim, the company will create a multiplexed MSIA assay that will be able to detect and profile five plasma proteins simultaneously, with a concentration range of 1 ng/mL to 1 mg/mL. Various ratios of antibodies will be immobilized in the affinity pipettes and the assay will



be tested on a standard solution of antigens in their physiological concentration to determine the best ratio of antibodies that produces comparable signals in the mass spectra for all proteins, without signal suppression issues. The limit of detection and CVs of the multiplex assay will be determined.

The overall objective of the phase II research is to develop a fully functional quantitative, automated, high-throughput, multiplex affinity protein capture technology platform and multiplex MSIA for analysis of low abundance cancer related proteins/peptides from bodily fluids. The company will start by identifying potential protein and peptide targets and their corresponding affinity reagents. Depending on the availability of such reagents, they will compile a list of more than 50 cancer-related proteins and peptides, and initiate development of individual MSIA for each target. The performance of the individual assays will be assessed in regards to their sensitivity and reproducibility. Those targets for which the assays pass the performance criteria will be evaluated for grouping into 10 sets of Multiplex MSIA, each targeting 5 protein/peptide analytes. The multiplex assays will also include an internal reference for quantification purposes. The multiplex assays, which represent the final product of this research, will be evaluated and validated in regards to their limits of detection, reproducibility, sensitivity, and specificity. As a final task of the project the company will assemble kits containing the multiplex assays, reagents, and protocols, and evaluate other product-related variables such as packaging and storage conditions.

#### **Accomplishments:**

During phase I period, Intrinsic Bioprobes Inc. has developed a MSIA for 5 plasma proteins with plasma concentrations in the 1 ng/mL to 1 mg/mL range, with inter- and intra-assay CVs < 10%. The company is preparing to start phase II of the project.

#### **2.5.4 Immunoaffinity Capture Couples with Ion Mobility**

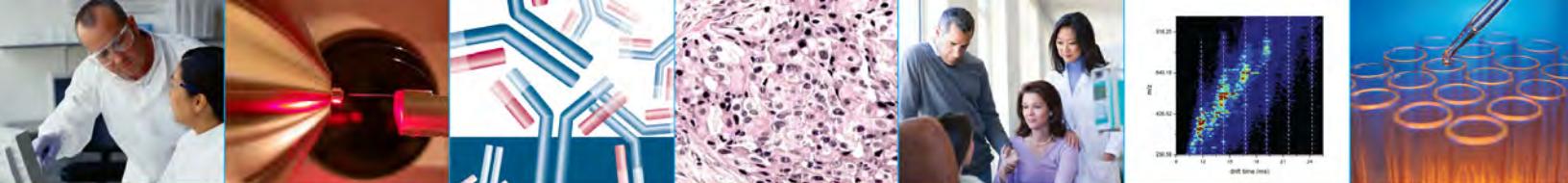
**Company: Predictive Physiology and Medicine, Inc. (PPM)**  
**SBIR**



**Principal Investigator:** Stephen J. Valentine Ph.D.

#### **Proposal: Immunoaffinity Capture Coupled with Ion Mobility Spectrometry (IMS)**

During the phase I period, PPM will couple immunoaffinity capture (IC) of prostate cancer biomarker candidate proteins with ion mobility spectrometry (IMS) - MS. The approach is unique in that the IC-IMS combination significantly reduces chemical noise thereby allowing greater utilization of the mass spectrometer. That is, higher sensitivity measurements will be achieved. There are three specific aims for the proposed work which include: 1) developing the prototype instrumentation and sample preparation



method; 2) validating the analytical platform; and 3) performing comparative proteomics profiling studies for a larger number of control and disease samples. The latter aim will be accomplished by developing an automated, high-throughput analytical platform and is directly focused on demonstrating the applicability of the technique for clinical analyses. The proteins that have been selected for immunoaffinity enrichment include putative biomarker candidates for prostate cancer diagnosis as well as the determination of disease progression and metastasis. Thus, the overall goal of this work is to not only produce an efficacious biomarker panel for prostate cancer but also develop instrumentation that can be used directly in the clinical laboratory. The company is planning to start the project in the near future.

### 2.5.5 Highest Sensitivity Cancer Marker Array on Quadraspec's Bio-CD Platform

**Company:** Quadraspec, Inc.  
**SBIR**

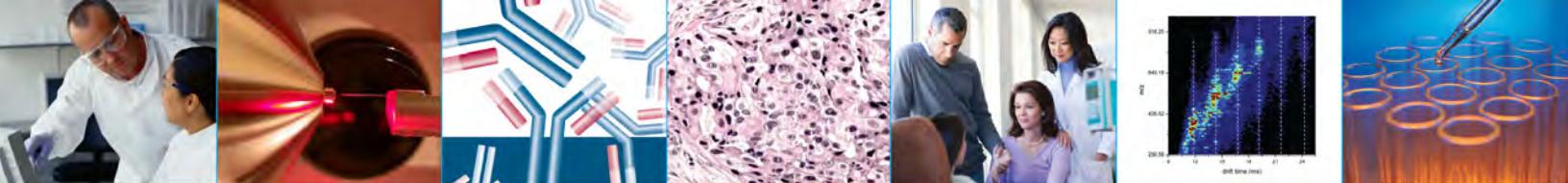


**Proposal:**

The company proposed to develop an array of 10 to 100 microfabricated disposable nano-immunoaffinity chromatography (NIAC) columns originating from a central 10 to 50  $\mu$ l well that can be used to selectively capture and enrich 10 to 100 cancer markers from bodily fluids as it flows into the column array from the central well where the sample is deposited. Quadraspec predicted that as sample flows into NIAC columns, this loading process will enrich and extend the dynamic range as well as enhance sensitivity by a million-fold while minimizing non-specificity. Detection is performed using conventional monitoring methods such as conjugated secondary antibodies and enzyme amplification. The goal was to develop a simple chip that can be used in clinical set ups.

**Set Milestones:**

- Create a small array of NIAC columns for concept validation that require less than 50  $\mu$ l of sample
- Enrich samples in NIAC columns to the level that limit of antigen detection will be < 5 pg/ $\mu$ l
- Demonstrate that NIAC columns can be used with a wide variety of antibodies
- Determine the degree of variability between different arrays (<10% variation)
- Determine that two antigens varying a million fold in concentration can be assayed with the same array
- Address the issue of cross-reactivity of antigens and non-specific binding
- An multiplex assay of at least 5 antigens would be achieved



### Accomplishments:

The physical construction of the channel system for performance of immunological assays was molded from poly dimethyl siloxane (PDMS). The NIAC column system was molded from PDMS using a metal mold. Two different coatings were tested in this work: The butyraldehyde coating and poly aspartamide. The latter appears to be the preferred choice. The data suggest that the contractor has reached a close to target goal of 5 pg/ $\mu$ l detection limit. Reproducibility of antibody immobilization in early prototypes exhibited a variability of 10% but poor flow reproducibility (>20% variability) could compromise quantification. It is suggested that electroosmotically driven flow (EOF) pumping that could be adopted in the future could minimize the variability to 5%. Potential issues related to cross-reactivity and non-specific binding were investigated using a number of examples. Contractors reached a dynamic range of  $10^5$ . Improved dynamic range may be achievable through capacitance-impedance sensor or possibly even with enzyme amplified assays as indicated by the contractor. Overall, significant progress was made toward achieving the goals of this project and milestones are substantiated with reasonable experimental data.

### 2.5.6 Automated Multi-Array Platform for Cancer Biomarkers

**Company: Meso Scale Diagnostics, LLC.**  
SBIR

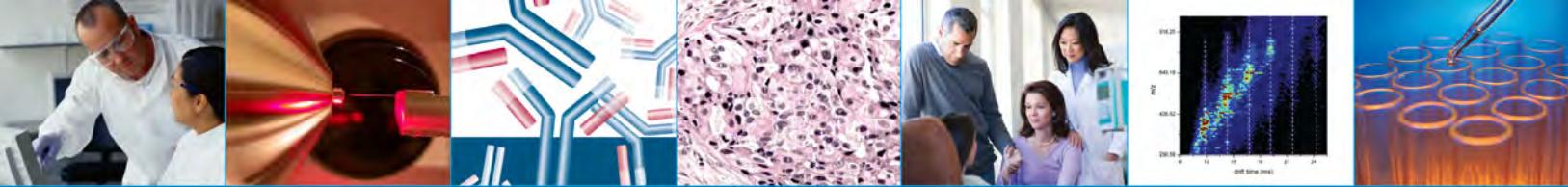


### Proposal:

The company proposed to develop a quantitative, automated, high-throughput, multiplexed immunoassay platform. This platform will be able to simultaneously measure up to 75 biomarkers using 30  $\mu$ l of body fluid samples. The features of this platform include ultrasensitive detection capabilities (< 1 pg/ml), a dynamic range of  $10^8$ , and a throughput of 60 to 300 samples per day. A consumable kit will be provided containing lyophilized detection antibodies, reconstitution assay diluents, and a 96-well plate containing the immobilized capture antibodies. It was projected that this platform will facilitate the measurements and verification of cancer diagnostic biomarkers.

### Set Milestones:

- Develop 4 new assays
- Transfer 3 assay panels to the 25-spot/well format
- Characterize assays and carry out method correlation tests for 5 analytes
- Conduct usability testing and evaluation of product prototype
- Make modifications to the prototype based on results obtained from usability testing
- Demonstrate dried reagent format
- Develop final specification for prototype development in Phase 2, and present finding to the NCI evaluation panel



#### **Accomplishments:**

Contractor developed 4 assays for the cancer related analytes, CA19-9, CA15-3, PSA and uPA with detection limits in pg- or sub-U /ml ranges. Contractor has also transferred three assay panels to the 25-spot/well format with minimal adjustments. They characterized the performance of each assay in terms of analytical sensitivity, dynamic range, precision, spike recovery and dilution linearity. They also compared the performance of 5 assays to 5 commercial tests. The results presented are generally acceptable or even exceeds criteria defined by CPTC program.

Contractor conducted usability testing and evaluation of product prototype by studying 50 patient samples. They were able to quantitate from 1.3 mg/mL SAA in one sepsis sample down to 0.2 pg/mL IL-1 $\beta$  in breast cancer patient samples. These results indicated that the product prototype has the ability to measure protein levels over an  $8.1 \times 10^9$  molar concentration range. The ability to achieve this result with 30  $\mu$ l of sample, with <13% CVs across 12 analytes demonstrates the potential that this product prototype may have in supporting clinical cancer biomarker discovery and validation.

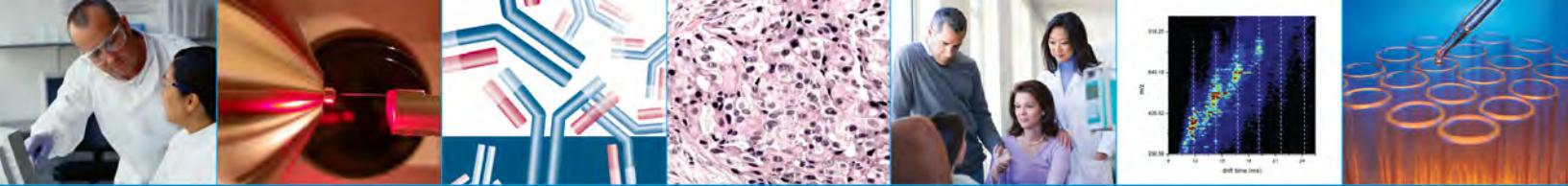
#### **2.5.7 Sensitive Protein Detection Combining Mass Spectrometry and PCR**

**Company: SEQUENOM, Inc.**  
**SBIR**

**SEQUENOM®**

#### **Proposal:**

The company proposed a novel approach to protein detection and quantification in bodily fluids that integrates three technologies with which Sequenom, Inc. and their collaborators have expertise in: immuno-PCR, competitive-PCR, and MS analysis using Sequenom's MassARRAY platform.



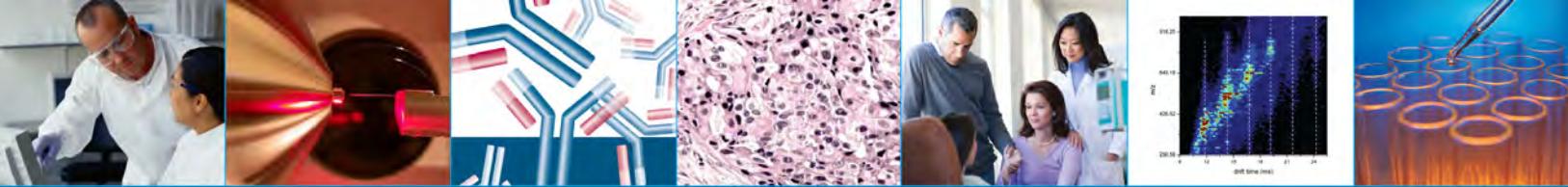
### Set Milestones:

- Protein selection
- Synthetic multiple nucleotide polymorphism design
- Verify detection of proteins using iPCR and MassARRAY
- Delineate quantitative boundaries in terms of sensitivity, dynamic range, reproducibility, precision, and specificity over multiple bodily fluids
- Repeat this characterization for at least 8 different protein/probe sets
- Build up to 8-plex using similar protein concentrations
- Verify wide dynamic range
- Multiplex using bodily fluids



### Accomplishments:

Sequenom has conducted proof of concept experiments with their collaborators using immuno-PCR, nucleic acid competitors, and the Sequenom MassARRAY platform to detect and quantify proteins. Tadpole and proximity probes were investigated for sensitivity and potential for multiplexing using six proteins. Sequenom selected 8 proteins for assay development and multiplexing purposes, some but not all of these proteins overlap with those chosen for proof of concept experiments. Probes for tadpole platform were manufactured and assessed. An 8-plex series of spike-and-recovery of antigen into buffer assays were performed. Sequenom also deployed a proximity ligation platform which resulted in high background when used in solution phase and so were not followed up. Instead the contractor decided to move forward with solid-phase proximity ligation and observed more favorable results in an 8-plex assay. This platform is not optimal in its current form as addition of a pre-amplification step has delayed its optimization. Sequenom also compared their tadpole data with ELISA platform for two analytes and indicated better dynamic range and sensitivity. Overall, Sequenom achieved considerable milestone in proof of concept and establishing an 8-plex assay for protein detection.



## 2.5.8 High-throughput Selection of Aptamers Against Cancer Biomarkers

**Company: Accacia International, Inc.**  
**SBIR**



**Proposal:**

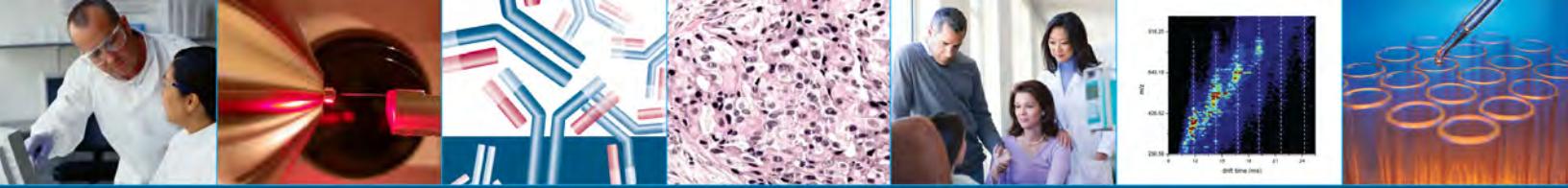
Aptamers (nucleic acid binding species) can be selected against a wide variety of analytes. Aptamers are not only interesting as affinity and diagnostic reagents but could prove to be useful as therapeutics. Accacia proposed to generate aptamer receptors and develop aptamer-based diagnostic assays such as the proximity ligation assay (PLA) that have much greater sensitivity than conventional ELISAs. Goals included selection of aptamers with low nanomolar Kds against ten targets (one aptamer/target), produce extended pools and generate high affinity “bivalent” aptamers. Paired aptamers that are superior to antibodies in both ELISAs and PLA assays will then be generated.

**Set Milestones:**

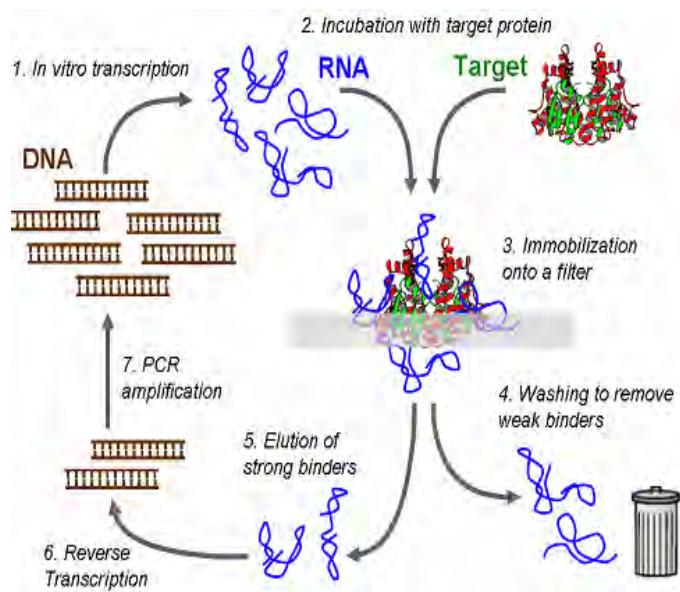
- Manually select aptamers with Kd values in the low nanomolar range against ten targets, which we will initially choose so as to best demonstrate the methods that will eventually be used with targets provided by the Clinical Proteomics Initiative.
- For the best 10 initial aptamers (one aptamer against each target protein), create ‘extended’ pools and use automated selection methods to generate ‘bivalent’ aptamers that bind to more than one epitope on a protein surface with very high (pM) affinities.

**Accomplishments:**

Accaccia has made selection against 3 targets: EGFR, ErbB2 and FGFR1.



**Figure 35. RNA Aptamer Selection to Protein Targets**

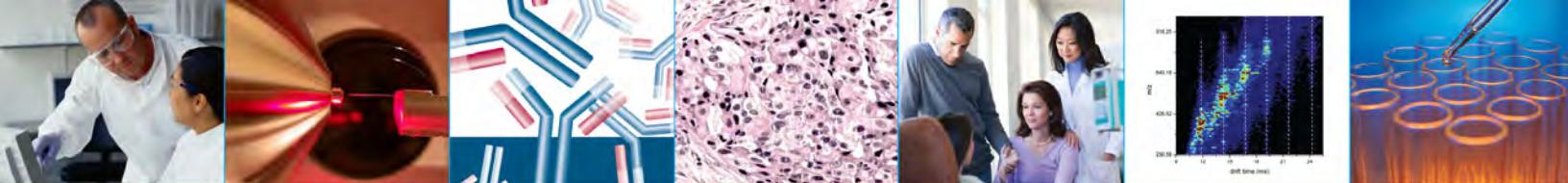


### 2.5.9 Proteomic Phosphopeptide Chip Technology for Protein Profiling

**Investigator:** Xiaolian Gao, Ph.D., University of Houston

#### Project Goals and Significance:

Dr. Gao is developing a proteomic phosphopeptide (PPEP) microchip technology platform that profiles proteins carrying phosphopeptide binding domains (PPBDs). Using profiles generated in laboratory experiments, along with predictive computational modeling (experimental data and PPEP/PPBD interaction database), ongoing work is attempting to demonstrate specific and quantitative measurements related to protein functions for proteins of biological importance. The methods developed will enable researchers to rapidly and vigorously develop peptide arrays for quantitative measurement of the proteins in biological systems or to use standard domain optimized peptide arrays to systematically profile biological samples for basic research or clinical importance. Over the long term, the methods will be used to establish domain-recognition systems for measurement of other types of domain-carrying proteins. The initial successes demonstrate the feasibility of proteomic protein profiling through phosphopeptide arrays. Current proteomic technologies identify proteins by either whole protein detection (2D electrophoresis and antibody binding) and/or peptide/phosphopeptide fragments (MS). This project fills the gaps between these technologies by detecting/profiling phosphoprotein binding domain proteins (includes



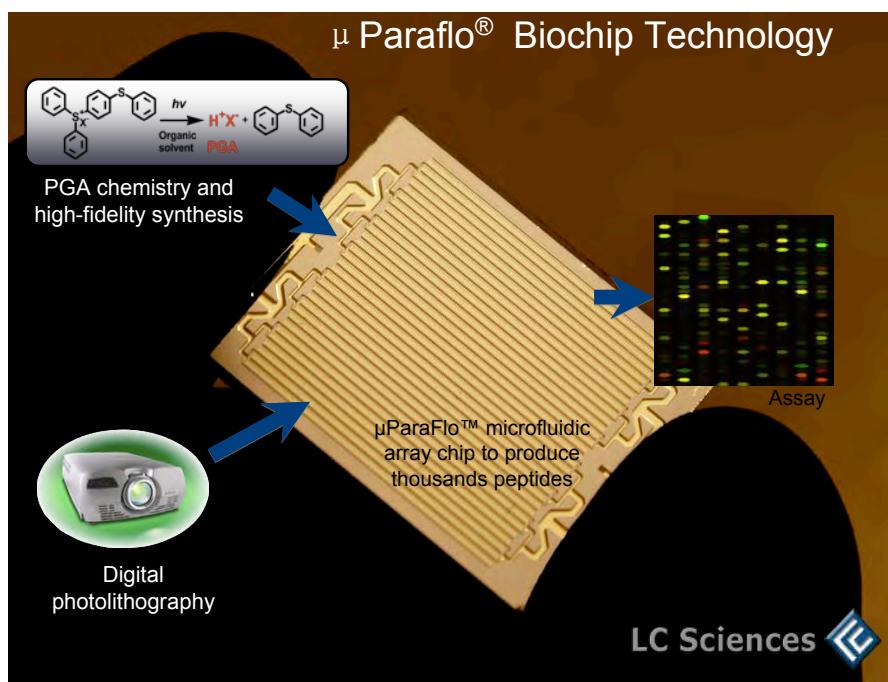
kinases, phosphatases and signaling proteins). These proteins are often active in cancer cells and are not easily detected by conventional proteomic methods.

### Accomplishments:

#### *Development of $\mu$ Paraflo® Biochip Technology*

The  $\mu$  Paraflo Biochip Technology (Figure 36) is an *in situ* synthesis using an activation agent generated using digital light and conventional chemistry onto microfluidic microchip. This unique  $\mu$ Paraflo system differs from conventional microarray in that it provides both a synthesis and assay platform. The use of these high density peptide chips for protein assay performance is unique. Current efforts are screening phosphoprotein binding proteins to develop clinical assays (diagnostic, prognostic and/or monitoring drug treatment).

**Figure 36.  $\mu$  Paraflo® Biochip Technology**



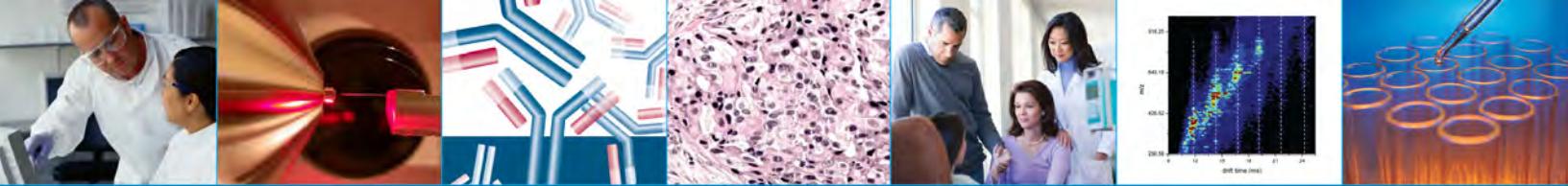
#### *Applying $\mu$ Paraflo® Biochip Technology*

##### o **Evaluation of pY peptides and SH2 Domains**

Hundreds of proteins active in cancer development processes contain SH2 domains. There are more than 100 different kinds of SH2 domains in humans. Initial studies screened “signature” peptides on the chip whose binding patterns could be indicative of their binding proteins *in vivo*.

##### o **Novel approach to determining Protein Binding Constants**

Current methods for determining protein binding affinities in a multiplex format use surface plasmon resonance (SPR). The  $\mu$ Paraflo peptide microarray system



is now capable of using a chip to measure 4000 proteins at once, with future capabilities as high as 30,000. This method has shown that the peptide microarray is a powerful application for quantitative and large scale comparative analysis of proteins.

***Developed a public/community database of human protein-protein interactions mediated by phosphoprotein binding domains (PepCyber)***

The PepCyber database focuses primarily on the interactions between binding domains in PPBPs and PPEP. A procedure has been established for database and literature citations to populate PepCyber with information regarding the interactions between PPBD and their PPEP substrates which are mediated through SH2 domain binding. A user accessible version of the PepCyber web site (<http://pepcyber.umn.edu/PepCyber>) has been created. The database currently spans 111 human SH2 proteins containing 121 SH2 domains. The phosphopeptide-protein interactions embedded in the database are divided into three categories which include established peptide-SH2 interactions, absence of a clear interaction but involvement of specific SH2 domain or SH2 protein.

***Developed a program package for designing & organizing peptide sequence patterns ( $\mu$ PepChip Pro)***

A preliminary “peptide array design wizard” to assist with the design of a peptide chip has been established. The URL of this web site is

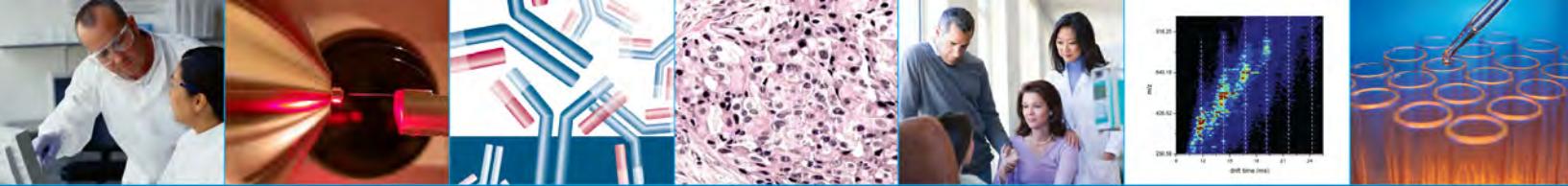
<http://micropepchip.umn.edu/microPepChip>. This web server performs a five step array design, which includes a function that searches in PepCyber for a set of “seed sequences” of interest, and a group of “sequence writer” functions that produce derived sequences from the “seed sequences” based on the users’ requirements. Presently, four writer functions have been implemented, capable of producing N-terminal truncated sequences, C-terminal truncated sequences, and Ala-scan sequences. Moreover, three types of control sequences, where pY is replaced by Ala, Phe, or Tyr, respectively, are available. The “peptide array design wizard” implemented in this version of the  $\mu$ PepChip Pro server is able to guide the user through the procedure of peptide array design which includes: choosing replicate numbers, selecting seed sequences, writing derived sequences based on the seed sequence, and writing the array layout into a downloadable Excel table.

**Next Steps:**

Complete the  $\mu$ PepChip Pro program package and demonstrate its utility. New methods and strategies will be developed to optimize specific binding of a target protein. The sample complexity will be increased from a single protein to complex mixtures (plasma, serum, tissue or cell lysate).

**Academic/Industry/Research Collaborations:**

University of Houston  
University of Minnesota  
Atactic Technologies



### 2.5.10 Aptamer-Based Proteomic Analysis for Cancer Signatures

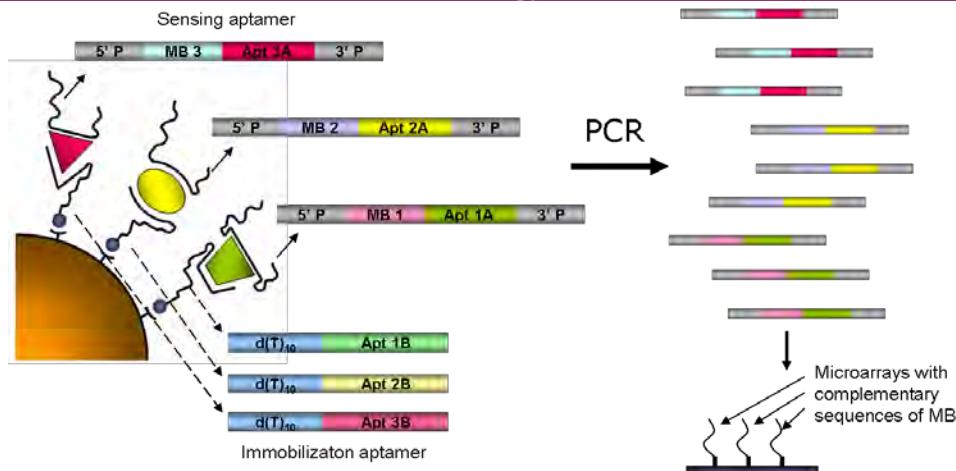
**Investigator:** Stephen P. Walton, Ph.D., Michigan State University

#### Project Goals and Significance:

Dr. Walton is developing an aptamer-based strategy for proteomics with *better sensitivity* and *dynamic range* than current array-based proteomic strategies. An aptamer is a protein capture reagent that utilizes an oligonucleotide tag as an identifier of the specific aptamer. This tag provides the option of ultimately using oligonucleotide microarrays to quantify proteomic signatures, which is advantageous from the standpoints of infrastructure and reproducibility.

With this cutting edge technology, Dr. Walton is developing a technique in which aptamers targeting different proteins are each labeled with a unique molecular barcode sequence (Figure 37). Molecular barcodes are a set of 20-nucleotide DNA/RNA sequences that serve as unique identifiers of specific aptamers. For each protein a pair of aptamers targeting the protein at distinct epitopes is used. One of the aptamers is immobilized on magnetic beads via the biotin-streptavidin interaction and serves to capture the protein. The sample is exposed to the aptamer pairs, followed by exposure to the streptavidin coated magnetic beads. After application of a magnetic field, the supernatant containing unbound material is removed. The sensing aptamers retained on the beads are amplified and measured by real-time PCR (1 to 4 proteins) or detected on an oligonucleotide microarray (tens to hundreds of proteins).

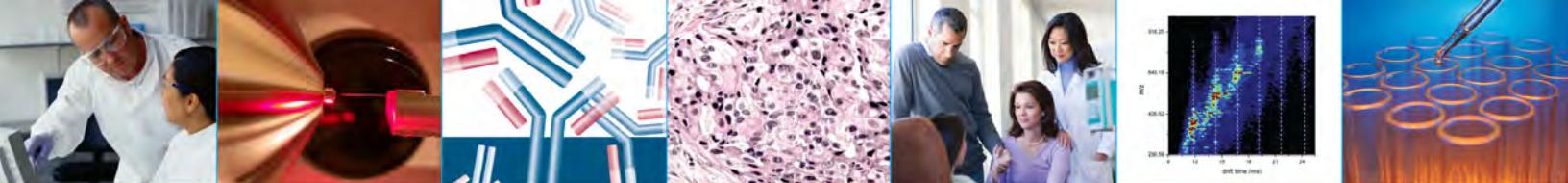
**Figure 37. Schematic of Proteomic Strategy**



#### Accomplishments:

**Developed aptamer assays for thrombin and PDGF-BB** Each assay shows high specificity for its target protein and emits a signal that correlates with its target protein's concentration. According to measurements from scintillation counting and PCR, each pair of aptamers has a lower detection limit of 50 nM.

#### Next Steps:



Selection of aptamers targeting haptoglobin, fibrinogen, and C-reactive protein. These proteins will provide a foundation for analysis of samples from cultured cells and sera. Dr. Walton will combine this strategy with other efforts using nucleic acid based parallel analytical technologies to further improve the technique and increase the number of target proteins.

#### Academic/Industry/Research Collaborations:

Andrew Ellington, University of Texas

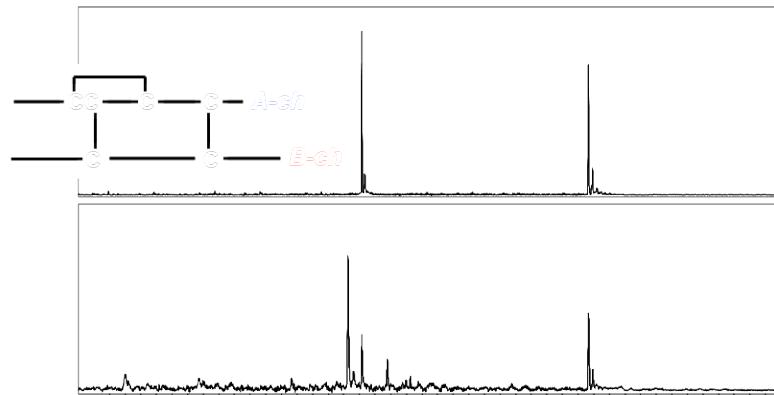
### 2.5.11 Top-Down Mass Spectrometry of Salivary Fluids for Cancer Assessment

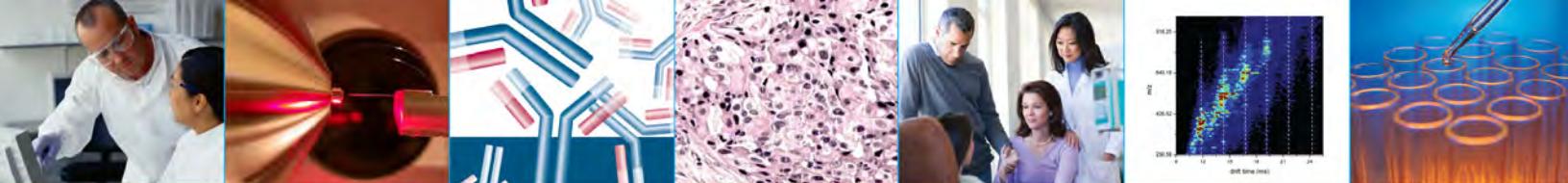
**Investigator:** Joseph Loo, Ph.D., University of California, Los Angeles

#### Project Goals and Significance:

Dr. Loo seeks to develop a new technology platform which combines laser desorption and electrospray ionization, termed electrospray assisted laser desorption/ionization (ELDI) and to provide top-down generated protein sequence tags to identify relevant cancer markers from biological fluids and tissues. ELDI works under ambient conditions for protein solutions deposited on a sample plate and produces multiply-charged ions. More accurate molecular weights of intact proteins can be measured for top down proteomic analysis, and the prospects of sequencing peptides and proteins using ELDI-MS/MS is promising. In addition, it was discovered that ELDI can potentially be combined with reactive protein chemistries (i.e., “Reactive-ELDI”). For example, disulfide bonds can be reduced by introducing the necessary reagents (e.g., DTT) by the laser pulse (if the analyte is in the ESI solution) or in the ESI solution (and the analyte is desorbed by the laser) (Figure 38). This offers the possibility of performing on-line reaction chemistries prior to MS measurements (e.g., proteolysis, chemical tagging, isotope labeling, etc), thereby greatly minimizing sample handling and processing prior to MS sequencing. The ability to combine on-line protein chemistries with separation technologies offers a unique platform for characterizing proteins, and potentially it can suggest new analytical advantages and strategies.

**Figure 38. Electrospray assisted laser desorption/ionization (ELDI)**





## Accomplishments:

### **Developed protein pre-fractionation strategies**

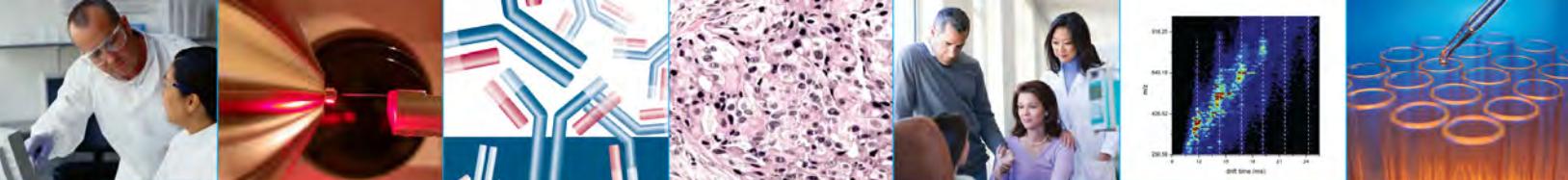
Protein pre-fractionation strategies developed to partition the salivary proteome for further characterization by top-down proteomics. The complexity of the proteome for their analysis by MS requires reduction and simplification to maximize the dynamic range of the analyses, i.e., identify lower abundance proteins from a “sea” of higher abundance proteins. Methods based on chromatography, isoelectric focusing, and affinity selection have been investigated as platforms for partitioning samples for MS characterization. This “divide and conquer” strategy has allowed us to identify many more proteins than previously reported. Our fractionation strategy continues to be focused on a protein-centric method, i.e., separation of intact proteins. Pre-fractionating by isoelectric point allows for larger protein loads without degrading gel electrophoresis resolution. Based on our experiments with human salivary fluids, the results demonstrate that it provides a convenient method to enrich for low abundance proteins. For the bottom-up method, by combining an additional stage of SCX fractionation with IEF prior to LC-MS/MS, we have expanded the catalogue of human salivary proteins to over 1500 (from approximately 1000 proteins previously).

### **Validation of ELDI MS Analysis**

Extended MS/MS top-down experiments have been shown to be compatible with ELDI for proteins, such as carbonic anhydrase (29 kDa), at high sensitivity (1 pmol). ELDI-MS shows a multiple charge distribution similar to that measured by conventional infusion ESI-MS. Top-down MS of bovine carbonic anhydrase (MW 29024.6) are carried out by CAD, and tandem MS data was collected up to MS<sup>4</sup>. MS<sup>2</sup> of the 29+ charged intact protein shows mostly multiply charged y-product ions, with the most abundant product ions  $y_{67}$  and  $y_{61}$  from the cleavage of the N-terminal amide bonds to Pro-193 and Pro-199, respectively (Figure 1B), which is consistent with the previously reported ESI-MS<sup>n</sup> data for carbonic anhydrase. MS<sup>3</sup> of the  $y_{67}^{7+}$  ion at 7598 Da shows y-products not observed from CAD of the intact protein. Further dissociation of the  $y_{61}^{6+}$  (MS<sup>4</sup>) at 7042 Da shows again mainly y-ions, including an abundant  $y_{47}^{4+}$  from cleavage of the Glu-212 / Pro-213 amide bond.

### **Improved Sensitivity**

ELDI-MS sensitivity for intact protein analysis is slightly higher when compared to DESI and MALDESI. For a medium sized protein such as 25.6 kDa chymotrypsinogen, Basile and coworkers reported a DESI detection limit (with a signal-to-noise of 3 or better) of 100 ng mm<sup>-2</sup> (*Anal. Chem.* 2007, 79, 3514-3518). MALDESI mass spectra from 0.8 μL of a 250 μM solution of 8.6 kDa ubiquitin deposited on the laser desorption surface have been reported (*Anal. Chem.* 2008, 80, 5266-5271). Using the ELDI platform, a mass spectrum was measured with a signal-to-noise of 3 for 29 kDa carbonic anhydrase from 10 μL of a 1 μM solution dried to a spot area of 13 mm<sup>2</sup>, corresponding to 23 ng mm<sup>-2</sup>.

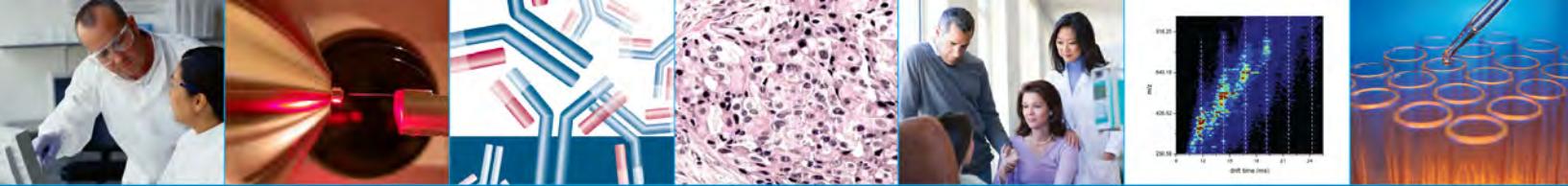


### **Next Steps:**

Future studies will continue to improve the sensitivity and applicability of the ELDI platform. Reactive ELDI allows one to perform ion/molecule reactions at the interface between the charged microdroplets and vapor phase of the laser desorbed material. Reactive ELDI provides an example of the fact that the ion chemistry traditionally performed in the gas phase at low pressure can be studied at the interface of microdroplets at atmospheric pressure. Additionally, development of an ELDI atmospheric pressure mass spectrometer source using a tunable mid-IR laser system based on optical parametric oscillator (OPO) technology is proposed. It is thought that mass analysis of a variety of samples could occur at room temperature and atmospheric pressure. Therefore, the sample would remain in its native environment and real time mass identification of sample constituents would be obtained with little to no sample pretreatment.

### **Academic/Industry/Research Collaborations:**

University of California at Los Angeles  
OPOTEK Inc.



## 2.6 Reagents and Resources Development

### 2.6.1 CPTC Reagents and Resources Core

Discussions with representatives from all parts of the cancer research community revealed a deep concern about the lack of access to affordable, well-characterized and validated affinity reagents and supporting resources. In order to drive the development of a central community core that would help accelerate biomarker discovery and validation, cancer diagnostics development, and therapeutics monitoring, NCI launched the Proteomic Reagents and Resources Core. This program within CPTC provides tools, reagents, enabling technologies, and other critical resources to support protein/peptide measurement and analysis efforts.

At the CPTC Annual Meeting in October 2008, the Reagents and Resources Core announced the launch of the Reagents Data Portal, a Web-based service created by NCI-Frederick in an effort to make reagents (i.e. antibodies), protocols/SOPs and characterization data produced within the CPTC program available to the scientific community (Figure 39).

**Figure 39. Examples of Reagents and Resources Available to Community**

#### Reagents

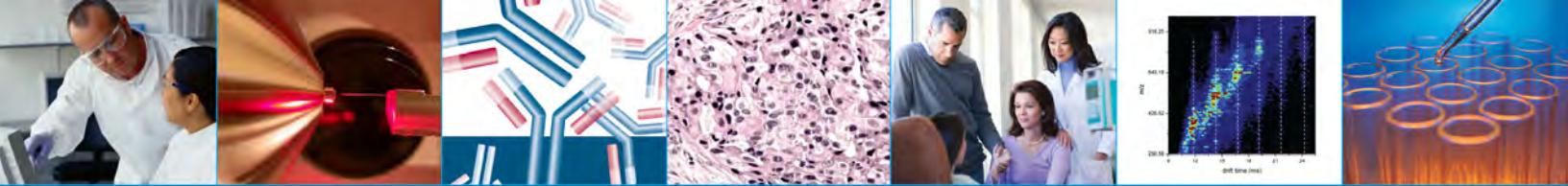
- Reference Materials (e.g., NCI-20, yeast proteome lysate)
- MRM Peptides
- MRM Standard Kits
- Antibodies
- Hybridomas
- Expression Vectors
- Mouse Biospecimens

#### Resources

- Tranche
  - CPTC
  - MPTI
- Standard Operating Procedures
- Software Packages
- FDA Documents ("Mock" 510(k))

#### 2.6.1.1 Reagents

The Reagents Data Portal is in the process of expansion as the initiative makes way for a great number of reagents in the pipeline that are needed for effective proteomic analysis. To date, more than 28 antigens and 81 monoclonal antibodies have been generated against human cancer-associated proteins and each antibody is added to the web portal once initial characterization (isotype, SDS-PAGE, Western Blot, and ELISA) has been generated.



### **Reference Materials, MRM Peptides and MRM Standard Kits**

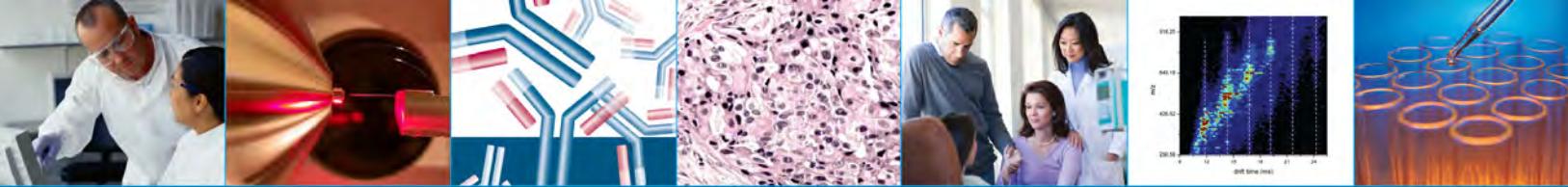
The goal of the NCI Clinical Proteomic Measurement Assessment Materials Program at the NIST is the development and documented characterization of complex biological mixtures to evaluate proteomic analysis platforms in the NCI CPTAC with the intention of developing proteomic SRM and to provide these developing materials and SRMs to the broader scientific community through the CPTC Reagents and Resources program.

NIST will prepare and characterize reference material including yeast, plasma, cell lysates which may be spiked with exogenous proteins or peptides. Additionally, reference materials may also include peptides and protein mixtures. These reference materials serve as important reagents in the development of measurement assessment for the CPTC and will support proteomic analysis by mass spectrometry and other analytical techniques such as affinity-based technologies. Such high quality, well characterized, and readily available protein mixtures and biological samples will be a key resource in evaluating proteomic technologies and will be essential for evaluating technology capabilities, standardizing workflow processes to reduce experimental variation, comparing experimental results, and consolidating qualified data from different experiments.

Heavy and light prototypic peptides will be generated and tested to serve as reference assays for quantitative proteomics.

### **Antigen Production**

A key component of the CPTC is the development of customized reference protein standards for mass spectrometry assessment studies. To this regard, ANL has established unique capabilities and expertise in the production, qualification, and characterization of peptides and proteins. This is an ongoing project, which connects the unique expertise of ANL in protein production with the needs of the NCI for the production of reagents as standards. At present, protein production at ANL serves a dual purpose for the scientific community. All ANL recombinant proteins are <sup>15</sup>N labeled for use in mass spectrometry studies within the CPTC program, while simultaneously used to generate monoclonal antibodies, thus providing affinity capture reagents for mass spectrometry and affinity array studies as well as for the scientific community at large. ANL supports the expression of the proteins through a variety of strategies using *E. coli* and yeast expression systems that employ a target analysis system to select protein targets based on informatics-based selection or target categorization (e.g. size, localization signals). The proteins are generated from cDNA clones, RNA, or PCR-based gene synthesis. The protein constructs are then purified and characterized by SDS-PAGE, and the solubility, yield and molecular weight determined. The final constructs are sequenced for accuracy. Each protein production process is carefully tracked and detailed for successful expression strategies.



### **Antibodies**

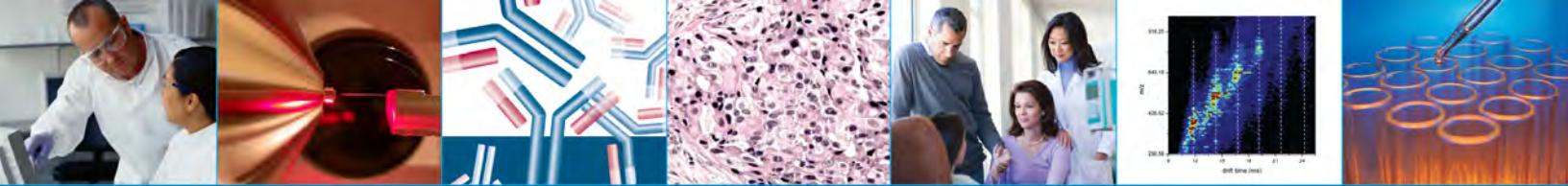
A key challenge for proteomic researchers is seeking out and acquiring high quality, well-characterized monoclonal antibodies. While numerous commercial reagent suppliers make antibodies available for research, their antibodies tend to be expensive and may or may not be extensively characterized. Thus, a researcher can at times be left guessing whether an antibody appropriate for their experimental platform is available for their studies.

Antibodies in the collection are being targeted against 1,261 tumor-associated proteins listed by Anderson and Polanski in 2006 as part of collaboration between CPTC and several laboratories and companies. ANL clones, expresses and purifies the proteins, which are then provided to private sector partners contracted for antibody generation through requests for proposals. Each contractor receives 40 proteins from ANL and generates 10 monoclonal IgG antibodies for each target protein, ultimately 3 of the 10 supernatants are purified and evaluated for in-depth characterization at 4 collaborating centers: NCI-Frederick, NCI's Center for Cancer Research Tissue Array Research Program (Bethesda, MD), the Harvard Institute of Proteomics (Cambridge, MA), and the Human Protein Atlas at KTH-Royal Institute of Technology (Stockholm, Sweden). Each antibody is analyzed using:

- ELISA
- Immunohistochemistry
- Immuno-mass spectrometry
- Nucleic acid programmable protein arrays
- SDS-PAGE
- Surface plasmon resonance
- Tissue arrays
- Western blot

All of the antibodies and their associated hybridomas are deposited with the Developmental Studies Hybridoma Bank (DSHB) at the University of Iowa; once characterization is complete, they are made available to the public at nominal cost.

<http://antibodies.cancer.gov>

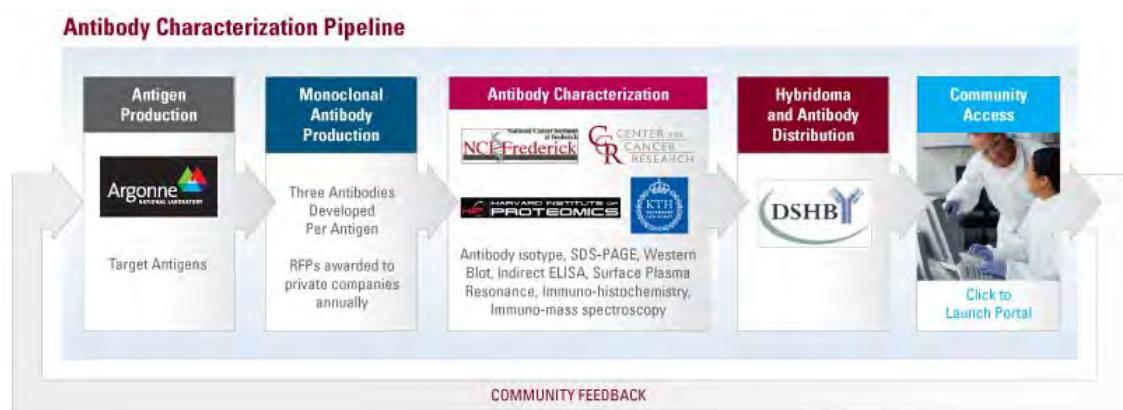


A key advantage of these antibodies is the depth of characterization data available. All of the associated characterization information is available for each antibody, as is detailed information on each protein (i.e., sequence, molecular weight, isoelectric point, alternate names, accession number(s), DNA source, and expression system). In addition, detailed standard operating procedures are posted for protein and antibody generation and characterization analyses.

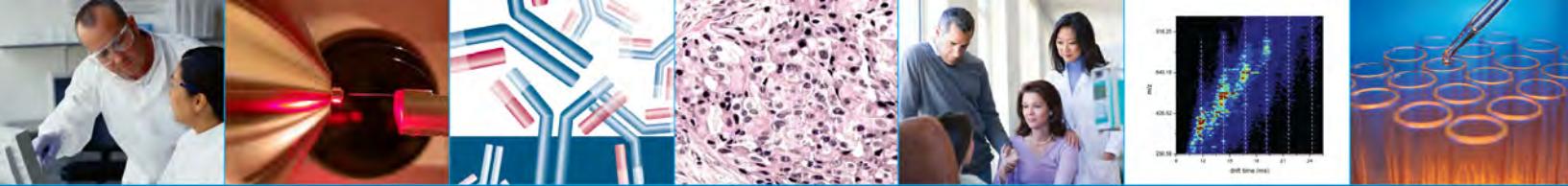
### **Antibody Characterization Pipeline: An International Collaborative Effort**

The goal of the Antibody Characterization Program of NCI's Clinical Proteomic Technologies for Cancer initiative (CPTC) is to have three monoclonal antibodies produced for each successfully expressed/purified recombinant antigen. Each antibody will be fully characterized (ELISA, Western, IHC, ImmunoMS, SPR and NAPPA) and made available to the scientific community (Figure 40)

**Figure 40. Antibody Characterization Pipeline**



The Harvard Institute of Proteomics (HIP) was founded to promote collaborative research and discoveries using novel high throughput technologies in the developing discipline of functional proteomics. Using the PlasmID repository, HIP has developed Nucleic Acid Programmable Protein Array (NAPPA), a unique high density protein microarray which is being applied in biomarker discovery for a broad range of disease areas. NAPPA evaluation for the antibodies generated within the CPTC program will be conducted in two phases. In the initial phase, the monoclonal antibodies will be screened against a subset of 1,000 proteins, of the full NAPPA antigens. Two arrays will be screened per antibody such that either each antibody is screened twice using the identical conditions, or each antibody is screened once using two different conditions. One monoclonal antibody against each successfully expressed recombinant antigen will be selected for further screening in Phase II. The optimal condition identified in Phase I



will be used to further characterize the selected monoclonal antibodies against the full complement of ca. 4500 antigens, which are printed on two NAPPA arrays in Phase II.

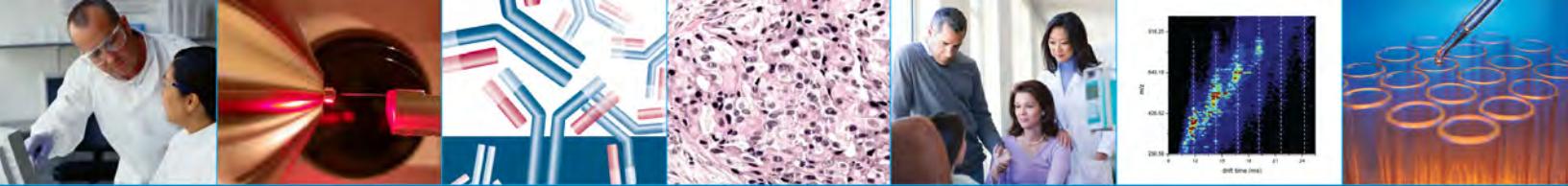
The Swedish Human Proteome Resource (HPR) program was set up to allow for a systematic exploration of the human proteome using antibody-based proteomics. The program hosts the Human Protein Atlas (HPA) portal with expression profiles of human proteins in tissues and cells. Protein profiling is performed using immunohistochemistry (IHC) on tissue and cell microarrays. The arrays comprise a large variety of normal human tissues, cancer tissues and human cell lines. Digital images are generated representing the staining and annotation of the respective images. As part of the IHC characterization, CPTC antibodies will be evaluated by HPA project in two phases. HPA will evaluate the antibodies in formalin-fixed and paraffin-embedded tissues according to their standard procedures. In the first phase, each antibody will be tested with tissue microarrays containing human tissues and cell lines using immunohistochemistry protocols defined by HPA. The outcome and results will be compared with published data, bioinformatics data and with results of other antibodies directed against the same target. A single antibody for each target will be selected for characterization in Phase 2. In Phase 2, microarrays for approximately 700 normal and cancer tissues and cell lines will be evaluated using the protocol defined in Phase I. All data, including the high-resolution images, will be made available to the CPTC community and listed in the Human Proteome Atlas.

### ***Reagent Distribution Partners***

The DNA clones which are used to create the recombinant proteins at ANL are deposited at Harvard University in the Plasmid Information Database (PlasmID) and are available to the scientific community. PlasmID currently contains over 85,000 plasmids, including collections produced by HIP, as well as plasmids deposited by individual researchers and programs such as CPTC. In addition to storage of these plasmids in a state of the art automated freezer system, a database is maintained which contains detailed annotation for each plasmid that can be accessed by users through the website. The goal is to collect plasmids, fully sequence validate them and then distribute worldwide.

The antibodies generated through the CPTC program are characterized through NCI-Frederick, and the monoclonal antibodies and hybridomas are deposited at the Developmental Studies Hybridoma Bank (DSHB) at the University of Iowa. The DSHB was created over 20 years ago by the NIH to bank and distribute hybridomas and the monoclonal antibodies (mAbs) they produce to the general scientific community. The main objectives are:

- **Low cost:** Monoclonal antibodies available through DSHB are much less expensive than commercially available counterparts. Researchers are thus able to test multiple antibodies, and continue to utilize those of interest without worry of expense.
- **Distribution:** To relieve scientists of the time and expense of distributing hybridomas and antibodies they had developed.



- **Maintenance:** To assure the scientific community that mAbs with limited demand would still remain available.

### **Mouse Biospecimens**

The NCI Mouse Proteomic Technologies Initiative, a component of NCI's CPTC, was designed to use animal models to develop and standardize technologies that help improve the accurate measurement of proteins and peptides linked to cancer processes. Launched in 2004, the program reflects a multidisciplinary and collaborative team approach to the development of standard tools and resources needed to accelerate protein biomarker discovery.

Mouse models of human cancer offer many distinct opportunities to optimize procedures for profiling major human cancers. The mouse serves as an *in vivo* model for cancer development that resembles the human model more closely than do cell lines and tissue samples. Mouse cells pass through many of the same physiological processes as human cells (such as apoptosis, angiogenesis, metastasis) during tumorogenesis. Researchers have previously elucidated many molecular pathways and frameworks of disease processes in mice, providing a framework that allows investigators to control many variables that could affect protein expression. By using the mouse model, researchers also gain the advantage of directly comparing aberrant protein levels measured in the mouse's cancer and surrounding tissues with those in the serum or plasma.

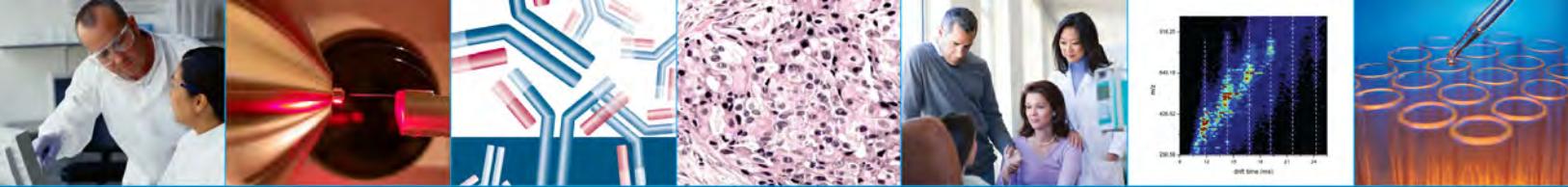
The NCI Biorepository of the Mouse Proteomics Technology Initiative exists to provide United States investigators with a resource of biospecimens from mouse models of cancer from a variety of cancer sites. Both plasma and tissue samples from these models were analyzed by proteomics labs at the Fred Hutchinson Cancer Research Center, the Institute for Systems Biology, and the Pacific Northwest National Laboratory. Plasma and tissue samples from 10 mouse models of cancer are available. References describing the methods and results of the generation and characterization of the mouse models are forthcoming. Because this is a finite resource, the NCI may not be able to satisfy all requests for samples.

**John M. Jessup, M.D.**  
**Chief, Diagnostics Evaluation Branch**  
**Cancer Diagnosis Program**  
**Division of Cancer Treatment and Diagnosis**  
**National Cancer Institute**

Proteomics is becoming increasingly important for personalized medicine in the post-genomic era because proteins are the catalysts for all biological processes and perform the instructions carried in the genetic code. As a result, new multiplex methods of measuring proteins, their post-translational modifications and

**DCTD**

Division of Cancer Treatment  
and Diagnosis



associated lipid and carbohydrate structures is becoming ever more important for personalized therapeutics.

The mission of the Cancer Diagnosis Program (CDP) within the NCI's Division of Cancer Treatment and Diagnosis (DCTD) is to improve the diagnosis and assessment of cancer by effectively moving new scientific advances into clinical practice. The program stimulates, coordinates, and funds resources and research on diagnostics and improved technologies to better characterize cancers in order to develop information that can aid cancer patients and their physicians in clinical decision-making. As such, there is a natural synergy between CDP and CPTC: CDP seeks to improve clinical decision-making today; CPTC seeks to develop the technologies and infrastructure to improve clinical decision-making in the near future.

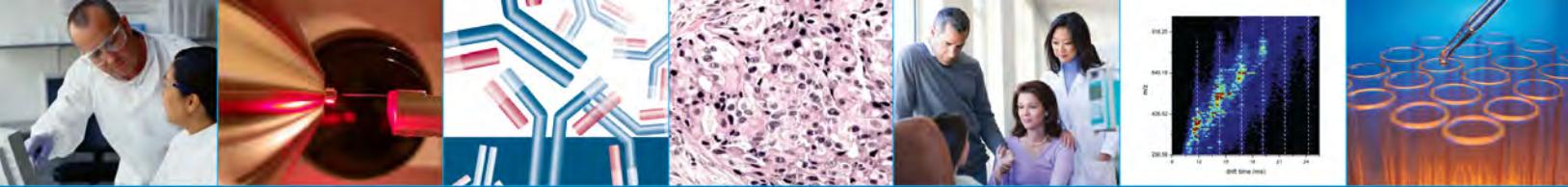
One year ago, CDP released a Request For Information (RFI) to an external community of approximately 1100 investigators asking what the community considered to be barriers to enhancing and hastening the incorporation of new molecular diagnostics into clinical practice ([NOT-CA-07-019](#)). The top two barriers were as follows:

1. Lack of specimens from patients with well-annotated clinical and pathologic histories
2. Lack of standards and reagents for clinical assays

Interestingly, the majority of investigators who were developing clinical assays were working with proteins; over 60% of these assays were multiplex. These clinically-oriented assays are still primarily using older technologies as ELISA and immunohistochemistry (IHC), indicating a need for standards and reagents for these technologies.

CDP informed investigators that CPTC will make available the peptides and antibodies that it creates. Approximately 200 separate requests were received for peptides that could be used as standards or for monoclonal antibodies that might be used to develop and validate assays for clinical research. In the first two sets of peptides and antibodies to be created by CPTC, a large fraction of the most commonly requested reagents are in the CPTC pipeline. Investigators are being notified as the peptides and antibodies become available and will be able to obtain these standards and reagents for research use at cost.

This is an outstanding example of how resources created by NCI for one purpose (to be standards for novel proteomics platforms) may be used by another program for a different purpose (to facilitate clinical assay development). This is so critical because the manufacturers of Research Use Only (RUO) materials are not required to standardize or certify the performance or quality of different batches of their reagents. In contrast to this "buyer beware" policy, the reagents supported by CPTC and distributed through CPTC-approved outlets will be characterized and certified for composition (peptides and antibodies) as well as activity and performance in different types of assays (antibodies).



This will be a tremendous aid for those investigators interested in developing assays for clinical research.

**Tim Harris, Ph.D.**

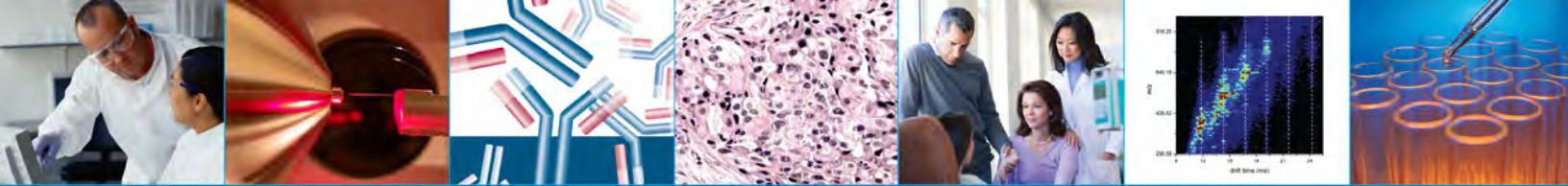
**Director, Advanced Technology Directorate  
SAIC-Frederick**

The critical need for well characterized affinity reagents for detecting proteins in cancer research has been well documented. In order to develop standardized reagents of this kind it is paramount that both starting materials and final affinity reagents undergo testing in a protocol that will be robust and reproducible. One of the best affinity reagents for detecting a variety of analytes are monoclonal antibodies. In order to accomplish the goal of providing the cancer community with standardized monoclonal antibodies the NCI is utilizing the resources of the Advanced Technology Program (ATP) at NCI-Frederick to develop methods and to characterize the antibodies such that a reproducible set of tools can be provided to the research community.



To this end an Antibody Characterization Laboratory (ACL) has been established. This laboratory is firstly the central recipient of materials. It performs the initial characterization of the protein antigens to which the antibodies will be raised, particularly as it relates to purity and it prepares them so that mice can be immunized to derive the monoclonal antibodies. The generation of the antibodies is done in collaboration with commercial partners selected through an RFP process. Candidate monoclonal antibodies are sent back to ACL and the final selection of at least 3 antibodies per antigen target is performed by the laboratory through evaluation of antibody binding studies using documented procedures, including ELISA, Western Blots and SPR studies. All assays are performed under standard conditions and SOPs are available so that each lab that is a recipient of the antibody can expect to see the same result using the same assay in their laboratory.

The literature is full of anecdotal studies where reproducibility cannot be obtained owing to poor antibody quality or characterization. One of the primary goals of this initiative is to make well characterized reagents available to the community. Once final antibodies are selected, a preparation of purified antibody is made for a final analysis that includes ELISA, Western Blot, SPR, antibody pairing studies, and immuno-mass spectrometry. All of these methods are being developed with the goal of procedural robustness and maximum information through data analysis.



The ACL also co-ordinates the distribution of antibodies for further analysis by reverse phase microarray analysis with the NCI-60 cell line performed through the Lab of Pathology, NCI; suitability for immuno-histochemistry by the Human Proteome Atlas in Sweden and binding studies against 4,500 selected proteins by the Harvard Proteome Institute. All data is collected by the ACL, reviewed and then entered into the web portal for public access (<http://antibodies.cancer.gov>).

The program to develop these reagents was initiated in the spring of 2008. To date, there have been 77 monoclonal antibodies generated against 19 antigens, and an additional 62 antigens are currently in production. The goal will be to have over 100 target proteins and 300 well-characterized antibodies by the end of 2009.

The goal of the ACL is to provide affinity reagents that are well characterized by a variety of techniques. Access to detailed characterization procedures will ensure that investigators have access to the best affinity reagents possible.

### 2.6.1.2 Resources

#### Tranche

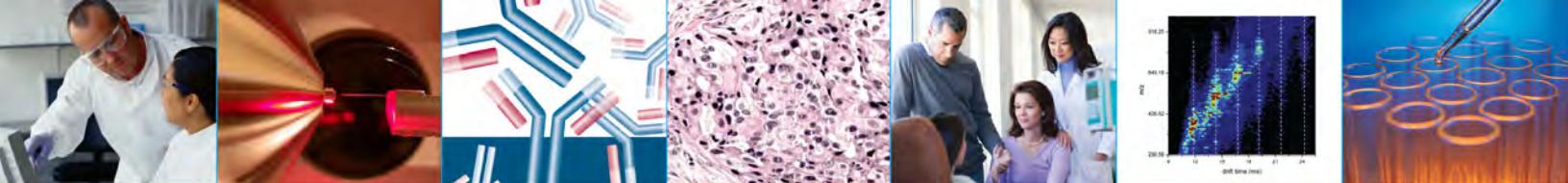
Tranche, an open source file sharing tool for scientific data, is the primary repository for proteomics data. Tranche allows collaborators to share data and research data regardless of file size, the number of files, or the file format. Tranche can also securely share data in pre-review before publication. Files can be encrypted and the passphrase is only available to the original uploader and whomever they share it with. In this way researchers can work in various locations all sharing data without worry about the security of the data being compromised. As such, Tranche is well-suited to serve as a data repository for mass spectrometry raw files.

Tranche has served as the repository for the CPTAC network, hosting all inter-laboratory data and metadata. In 2009, Tranche and its associated annotation tool are in the process of becoming caBIG™-compliant, making CPTC data accessible to the broader biomedical research community.

The Mouse Proteomic Technologies Initiative (MPTI) collected tissue and serum measurements from mouse models of different types of cancers using analytical techniques such as MS. Nearly 1 terabyte of MPTI raw data has been deposited into Tranche and is shared with the scientific community.

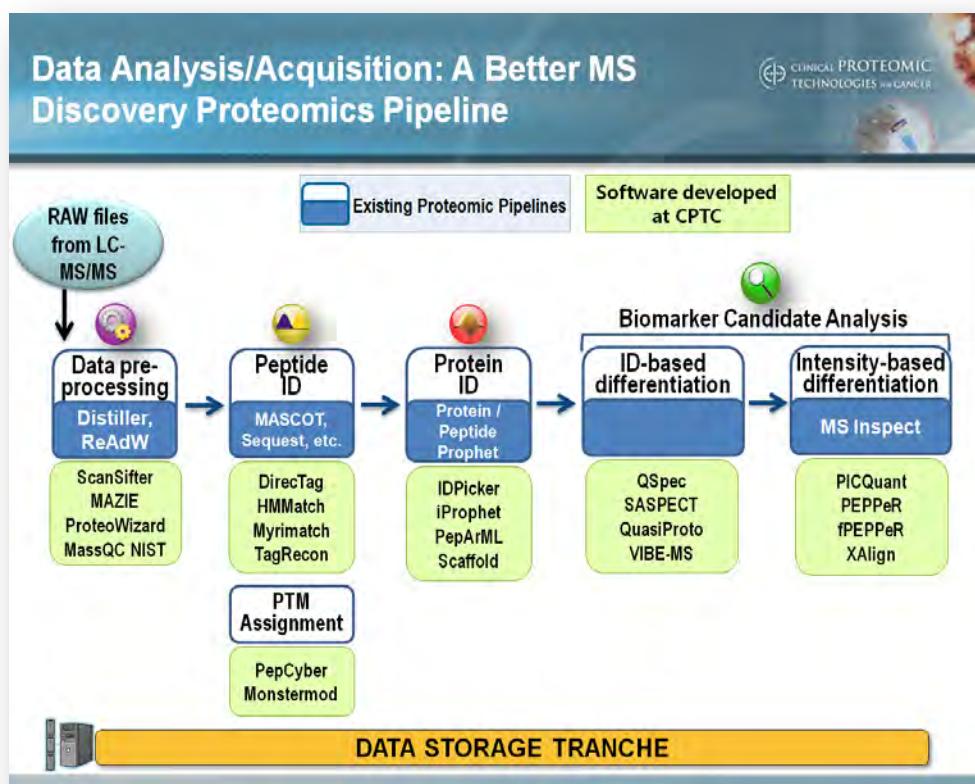
#### Standard Operating Procedures

A biospecimen collection protocol is the first multi-center SOP generated within CPTC and is the result of consolidated protocols from five leading clinics and associated



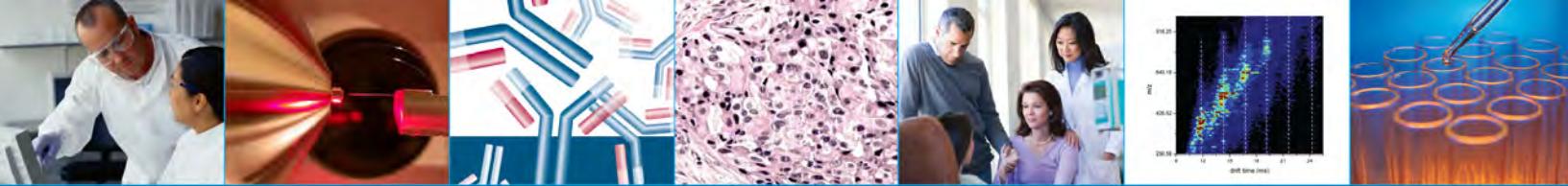
collaborators. CPTC worked closely with OBBR to develop this protocol. A Biospecimen working group, consisting of at least one investigator from each of the five network centers developed the framework for the SOP, including need for a high-quality biospecimen cohort, collection, processing and storage. Through a centralized biorepository, each center has access to a greater number of high-quality samples to evaluate.

### Software Packages:



### Data Pre-processing

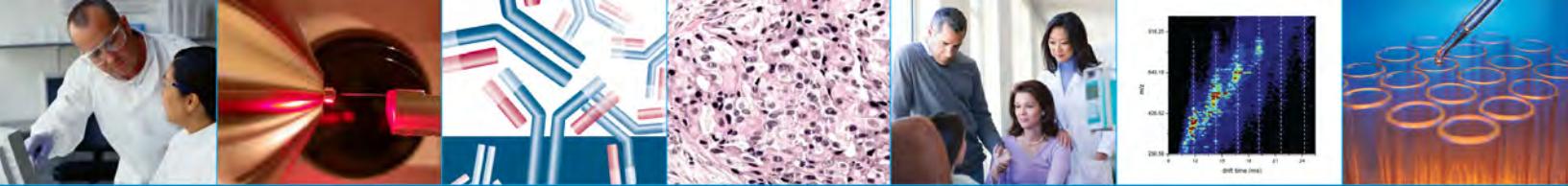
- ScanSifter: The “ScanSifter” algorithm, a Vanderbilt in-house-developed software, assesses the quality of each raw spectrum and discards poor quality spectra. As such, this application streamlines data analysis systems. In detail, the algorithm reads tandem mass spectra stored as centroided peak lists from Thermo RAW files and encodes them to text-based files. Spectra that contain fewer than six peaks or that have less than 20 measured peaks in the total ion chromatogram are not encoded. If 90% of the intensity of a tandem mass spectrum appears at a lower  $m/z$  than the precursor ion, a single precursor charge is assumed; otherwise, the spectrum is processed under both double and triple precursor charge assumptions.



- MAZIE (Mass and Charge (Z) Interface Engine): This software improved identification of peptide ion mass and charge, based on the isotopic distribution of peptide ion envelopes. This software will be distributed freely to the research community upon publication. MAZIE was written to enhance the fidelity of ion clustering. MAZIE is however suitable for general use to enhance database searching for mid-resolution mass spectrometers.
- ProteoWizard: The ProteoWizard software provides modular and open source, cross platform tools and libraries. The tools perform proteomic data analysis, while the libraries enable rapid tool creation by providing a robust, pluggable framework that simplifies and unifies data file access and performs standard proteomics and LCMS dataset computations.
- MassQC: Mass QC is a software package that serves to diagnose MS instrument hardware. Using data from CPTAC inter-lab studies, NIST developed a number of metrics to assess instrument performance. Through careful examination, NIST developed relationships between specific metrics and aspects of the measurement process. For instance, a decrease in the chromatographic elution time for a sample may indicate that a column should be replaced. ProteomeSoftware, a small software company in Portland, Oregon, built a graphical user interface over the NIST metrics. The resulting software package is called MassQC and was released June 5, 2009.

### Peptide ID

- DirecTag: DirecTag, is a tag-based identification which has been shown to be an accurate way to recover sequences from tandem mass spectra. It uses identification of peptides through sequence tagging using automated sequence tag inference. The algorithm has been evaluated on a diversity of MS instruments, from TOF/TOF to quadrupole ion trap. DirecTag has been released with source code to the research community.
- HMMatch: HMMatch tool demonstrates the ability to confidently assign more peptide identifications than is possible with a single search engine score, with no loss of statistical significance. The increased number of peptide identifications improves protein coverage and the ability to discern protein isoforms. A statistical significance model permits HMMatch scores to be compared with each other, and with other peptide identification tools, on a unified scale. It is possible to extrapolate HMMatch models beyond a single peptide's training spectra to the spectra of related peptides, expanding the application of spectral matching techniques beyond the set of peptides previously observed. In contrast to sequence-based tandem MS search engines used for peptides, spectral matching can make use of the intensities of fragment peaks in library spectra to assess the quality of a match.
- Myrimatch: MyriMatch makes more effective use of fragment ion intensity in comparison to X!Tandem Expect and Sequest XCorr and is robust against noise peaks. MyriMatch has been selected as the standard search engine for processing the data sets of the CPTAC unbiased WG.
- TagRecon: The process of tag reconciliation can allow amino acid changes to either side of the inferred sequence. In this process the tag sequences for an



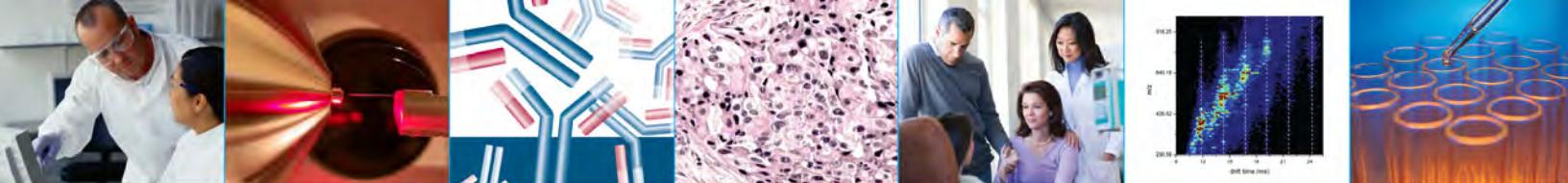
MS/MS are reconciled against the protein sequences from the database. The “TagRecon” software conducts this process using the same scoring algorithm as in MyriMatch. Therefore combining TagRecon and MyriMatch search results increases confident peptide identification.

### PTM Assignment

- PepCyber: The PepCyber database focuses primarily on the interactions between binding domains in PPBPs and PPEP. A procedure has been established for database and literature curation to populate PepCyber with information regarding the interactions between PPBD and their PPEP substrates which are mediated through SH2 domain binding.
- Monstermod: MonsterMod, a successor to the P-mod algorithm, matches a user-supplied list of peptide or protein sequences to a collection of tandem mass spectra via the MVH scorer (first developed for MyriMatch). This functionality will be separated into the "MassRecon" data analytical tool and a PTM explorer tool for visualizing and interacting with the resulting identifications.

### Protein ID

- IDPicker: The IDPicker tool enables users to organize experimental data into complex hierarchies. It was developed for protein assembly and has proven to be invaluable in generating tables of spectral counts that can be used for identifying candidate biomarkers in large cancer data sets. It has also been instrumental in organizing the complex datasets from the Unbiased Discovery inter-laboratory studies.
- iProphet: iProphet, allows more precise integration of information supporting the identification of each unique peptide sequence from multiple MS/MS spectra. iProphet takes as input PeptideProphet spectrum-level results from multiple LC-MS/MS runs, and then computes a new probability at the level of unique peptide sequence. The new framework allows combining results from multiple search tools, and also takes into account other supporting factors including: number of sibling experiments identifying same peptide ions, number of replicate ion identifications, sibling ions, and sibling modification states.
- PepArML: The PepArML Meta-Search engine provides access to large scale MS/MS sequence database searching infrastructure to researchers and labs without the computational resources or personnel to implement a distributed computing strategy in-house. Furthermore, this infrastructure provides a mechanism for search engines not intended or designed to run in a distributed or parallel fashion to be used in a distributed environment, without the need to modify the individual search engines — eliminating the need for the many ad hoc distributed computing solutions embedded in each individual search engine. Lastly, the meta-search engine is designed to be self-contained, platform independent, and require minimal operating system support, making it suitable for installation in small to medium size labs with little distributed computing expertise.
- Scaffold: Scaffold is a computer program that integrates search results from three algorithms (Sequest, X! tandem and Mascot) to generate peptide



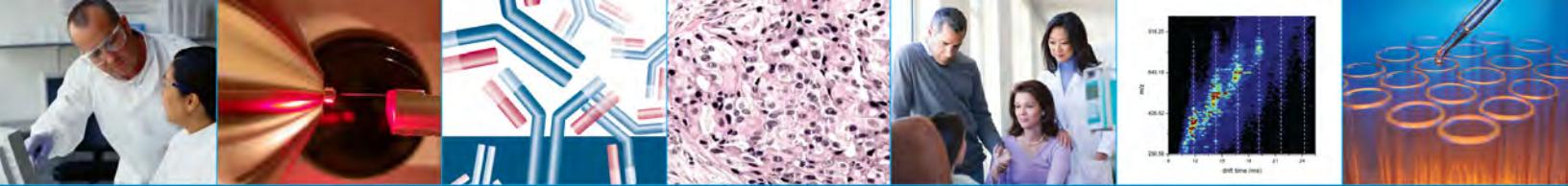
identification and protein identification probabilities. Scaffold also displays an overview of the protein identifications that can be validated by probability scores. Protein information can also be used to detect false positives and also examine the peptide and spectral evidence used for identification.

### ID-based Differentiation

- QSpec: The QSpec method is for data generated by the spectral count method. The spectral count method has become an accepted method for label-free quantitation in proteomics and comparable to measurements of extracted ion-chromatographic intensities. The spectral count method has the advantage of being applicable towards shotgun proteomics data using medium or even low resolution mass spectrometers, but comparison of proteins between two conditions is restricted to only those that are identified by MS/MS scans in both conditions.
- SASPECT: SASPECT provides a function for identifying differentially expressed proteins between two sample groups using spectral counts from LC-MS/MS experiments. SASPECT employs the commonly used “spectral-count” assumption: the probability of a protein’s being observed in one LC-MS/MS experiment is proportional to its abundance in the complex sample. However, in contrast to spectral counting, SASPECT uses the Boolean values of whether the spectral count is greater than zero instead of the raw values of spectral counts, for the latter are more subjected to the changes of various experimental factors. In addition, by properly controlling the FDR, SASPECT provides quantitative guidance in peptides and proteins selection.
- QuasiProto: QuasiProto is designed for spectral count differentiation in complex proteomic data sets. The software works from IDPicker tables that report the numbers of spectra matched to each protein. QuasiProto computes q-values for proteins by means of a quasi-likelihood model based on these spectral counts. This model enables the incorporation of features such as differences in instrument performance over time.
- VIBE-MS: The recently developed VIBE Toolkit for *Mass Spectrometry* gives users access to an integrated, modular environment for mass spectrometry data classification. The software provides an extensible ‘drag-and-drop’ graphical interface for creating workflows, which is an ideal environment to efficiently evaluate and optimize MS analysis pipelines. The software provides the required flexibility in the selection, comparison, and optimization of these analysis methods, as well as the optimization of the entire analysis pipeline. The ease with which modules can be removed from a pipeline, replaced with an alternate module, re-parameterized and used reduces the amount of time and effort required for a researcher to arrive at the optimal analysis protocol for a given problem.

### Intensity-based Differentiation

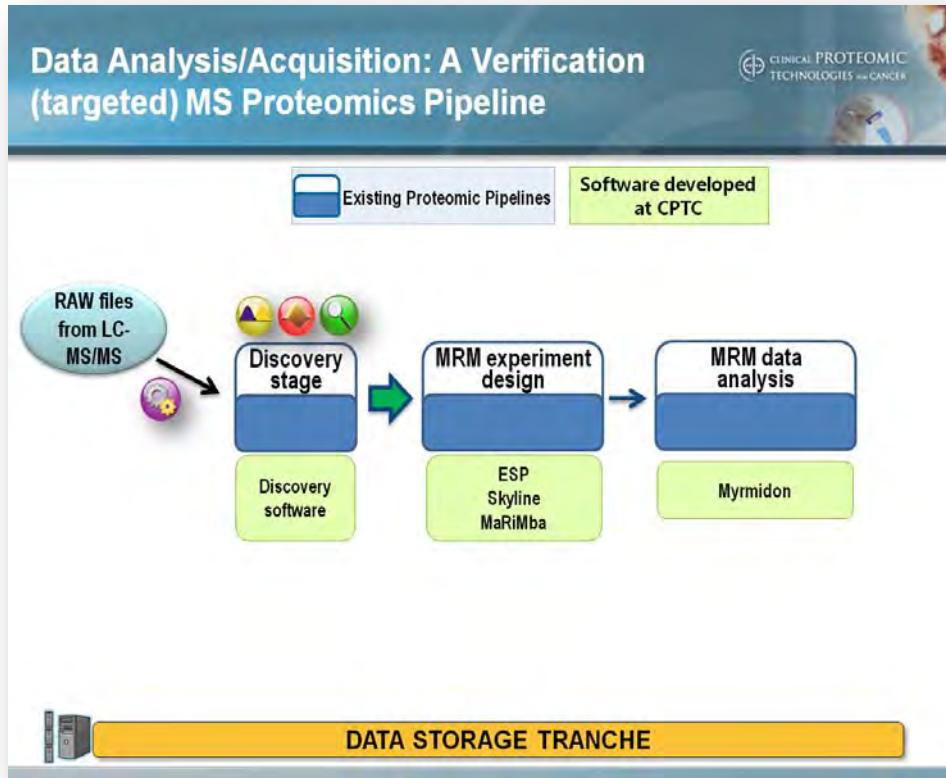
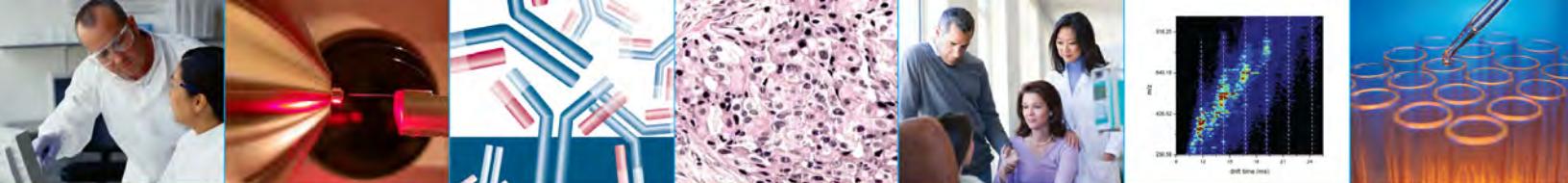
- PICquant: A new stable isotope mass tag ( $^{13}\text{C}$  Phenylisocyanate [PIC]) that has several advantages for protein quantification in complex mixtures was developed. This platform includes: 1) custom designed software (PICquant) to automatically quantify labeled peaks, 2) a spectrum-comparison algorithm that groups spectra



into a registry of spectra representing unique peptide families, and 3) enhanced peptide sequencing by distinguishing b- and y-ion series in CID spectra.

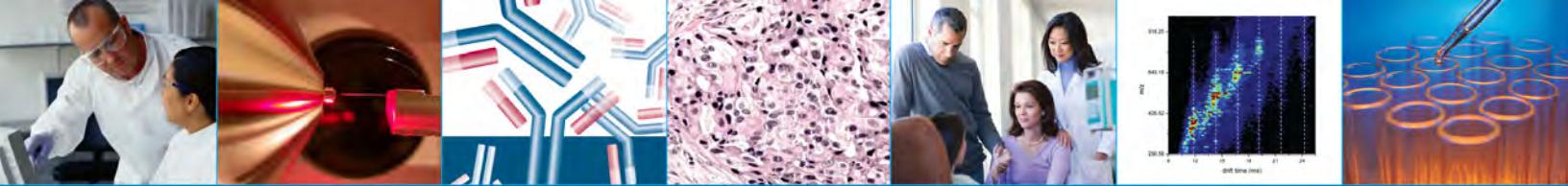
Completing the platform is a clinical registry that links acquired specimens to current and prospective clinical information including outcomes and that enables multivariate clustering of disease states with quantified protein families. The mature PICquant platform will provide nearly completely automated data analysis, allowing assembly of numerous patient samples into complete protein abundance profiles akin to gene expression array data. Additionally, the expanding Registry database will be of use to other investigators processing either PIC-labeled or unlabeled peptide spectra.

- **PEPPeR:** A Platform for Experimental Proteomic Pattern Recognition, “PEPPeR” uses high resolution and high mass accuracy LC-MS data from state-of-the-art mass spectrometers, and appropriately combines pattern-based (unidentified peptide peaks) and identity-based (peptides sequenced via MS/MS) information to generate peptide quantitation—thereby extending biomarker discovery to all charge identified MS1 peaks. From a computational perspective, the uniqueness of this approach arises from the use of: (i) identified peptides to guide alignment of unidentified peaks; (ii) matching unidentified peaks across multiple samples using mixture model based peak matching; and (iii) adaptive matching tolerances automatically calculated for each experiment.
- **fPEPPeR:** A recently-developed extension of PEPPeR, named fPEPPeR, incorporates the very first methodology for processing and computationally reassembling peptide fractions from multidimensional fractionation to facilitate data analysis at the sample level. The method works well despite imprecision of fraction boundaries or other variations during fractionation. In fPEPPeR, the PEPPeR peak-matching algorithm has been adapted to identify the same peptide species (peak) not only across multiple samples, but also across different fractions. fPEPPeR outputs intensity measurements for a common set of peaks spanning all the fractions under consideration. A sample is then computationally reassembled (i.e., defractionated) by summing the intensity measurements for each matched peak across all fractions from that sample. The defractionated samples can then be subject to biomarker discovery, class prediction, clustering and other pattern recognition algorithms without regard to the fractionation or any variations therein. This software is freely available as a GenePattern module.
- **XAlign:** XAlign is a two-step alignment algorithm. The first step is to detect significant peaks that are common to all samples. In the second step, all samples are aligned to the median sample using refined *m/z* and retention time variation values, where pattern recognition is applied as needed.



### MRM Experiment Design

- ESP: The first step in developing effective MRM assays is to identify the appropriate peptides. Identification of the most effective signature peptides, particularly in the absence of experimental data, remains a major resource constraint in developing targeted MS-based assays. ESP is a computational method that uses protein physicochemical properties to select high-responding peptides and demonstrate its utility in identifying signature peptides in plasma, a complex proteome with a wide-range of protein concentrations. This method, which employs a Random Forest classifier, facilitates the development of targeted MS-based assays for biomarker verification or any application where protein levels need to be measured.
- Skyline: Skyline is a Windows client application for building Selected Reaction Monitoring (SRM) methods and analyzing the resulting mass spectrometer data. It aims to employ cutting-edge technologies for creating and iteratively refining SRM methods for large-scale proteomics studies. The core focus of v0.5 is analysis of result data, building on the successful method creation features of v0.2. The library spectrum can then be used to identify fragment ion peaks ranked by intensity, and enable the user to define how many product ions ( $n = 3$ ,



4, 5, and 6 in this comparison) are required to provide a specific and selective measurement given the target sample.

- **MaRiMba:** The MaRiMba software tool automates the creation of explicitly defined MRM transition lists required to program triple quadrupole mass spectrometers in such analyses. MaRiMba creates the MRM transition lists from pre-existing or custom spectral libraries, restricts output to specified proteins or peptides, and filters based on precursor peptide and product ion properties. MaRiMba can also create MRM lists containing corresponding transitions for isotopically heavy peptides, for which the precursor and product ions are adjusted according to user specifications. This open source application is operated through a graphical user interface incorporated into the Trans-Proteomic Pipeline (TPP), and it outputs the final MRM list to a text file for upload to MS instruments.

### MRM Data Analysis

- **Myrmidon:** Myrmidon streamlines the automated analysis of large-scale SRM experiments. The software is compatible with files from several manufacturers of triple quadropole instruments and can make use of SRM methods defined in Skyline software. Scaling across multiple replicates in multiple cohorts, the software assists in the recognition and integration of chromatographic peaks, producing reports suitable for statistical interpretation.

### FDA Documents – “Mock 510(k)”

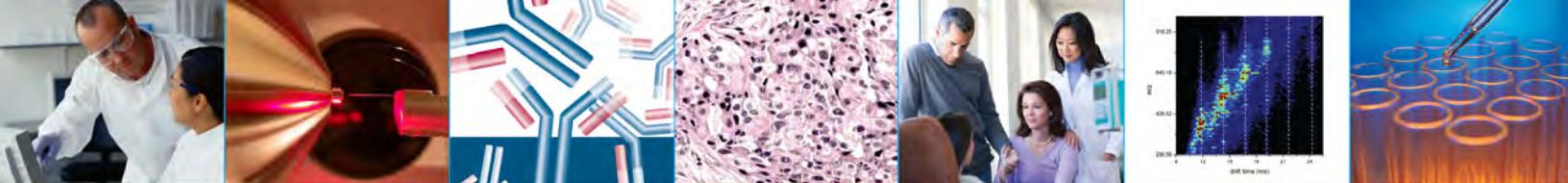
The NCI-FDA Interagency Oncology Task Force (IOTF) was established in Spring 2007 and a Molecular Diagnostics Workshop was held in Fall 2008. Mock 510(k) documents were proposed and are currently being drafted; one for proteomic MS platforms and the other for multiplex affinity platforms. This mock 510(k) documents will assist the FDA in understanding the role mass spectrometry can play in clinical assays, while serving as a reference guide to the proteomic community.

## 2.6.2 Global Production of Disease Specific Monoclonal Antibodies

**Investigator:** Barry L. Karger, Ph.D. Northeastern University

### Project Goals and Significance:

Dr. Karger has demonstrated the success of a method for generating and screening disease specific mAbs to native glycoproteins in plasma. In this study, this method was used to discover mAbs for biomarkers in lung cancer. Mice were immunized with a preparation of native glycoproteins isolated from the plasma of patients with non-small cell lung cancer (NSCLC). A method was established to produce the immunogens, where the concentrations of the major plasma proteins were reduced using a multi-step depletion protocol to produce depleted/fractionated plasma for the immunization protocol. A high throughput platform for large scale production of mAbs that resulted in a library of mAbs which discriminate lung cancer and matched controls was established.



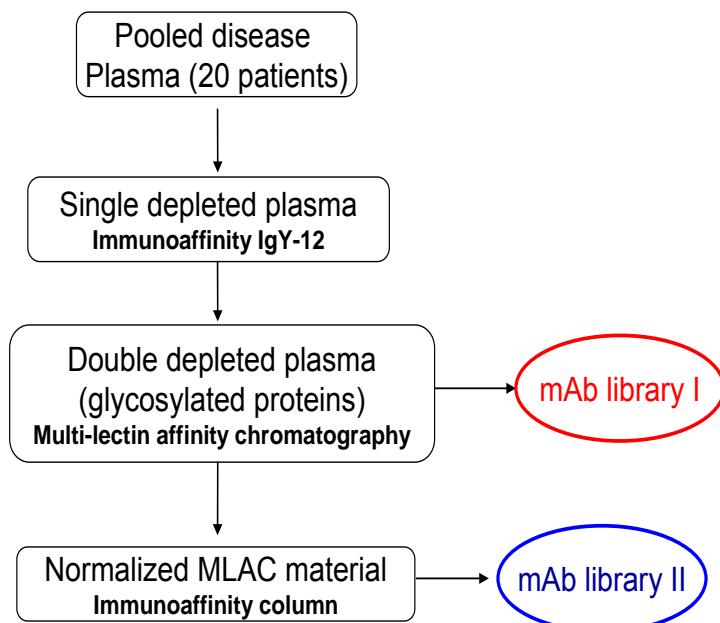
This platform led to the discovery of high quality, well-characterized lung cancer specific mAbs. Unlike other methods based on recombinant proteins or peptide conjugates, this method can directly produce native form-specific immuno-reagents for biomarker studies. Further evaluation/characterization of mAbs produced in this study could serve to complement the Reagents and Resources component of CPTC.

### Accomplishments:

#### *Establishment of NSCLC library*

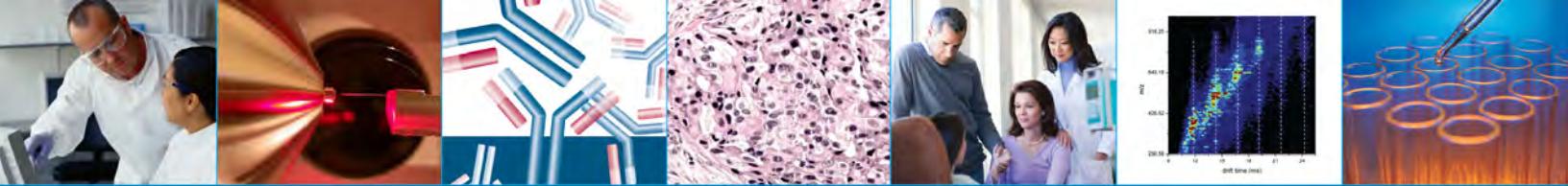
Plasma from 20 patients diagnosed with NSCLC and 20 clinically matched normal healthy controls were obtained. An equal amount of the plasma was pooled and subjected to successive depletion steps aimed at reducing the level of the most abundant proteins thereby reducing the large dynamic range of protein concentration. Three different steps of depletion were applied (Figure 41). Mice were immunized with fractionated native plasma glycoproteins, subsequent hybridoma cell lines which produced sufficient amounts of IgG (>50 ng/ml) were selected (total of 1051). A total of 184 positive hits (17%) were shown to discriminate between lung cancer and matched normal controls.

**Figure 41. Multi-step sample preparation for immunogen preparation**



#### *Antigen Identification*

Mice were immunized with a complex mixture of unknown glycoproteins to produce the mAbs. The goal was to identify the target antigens which corresponded to the mAbs. To



determine the identity of the lung cancer specific antigens a workflow which combined immunoprecipitation, SDS-PAGE, Western Blot and LC-MS/MS was utilized. To date three of the antigens corresponding to mAbs have been identified ( $\alpha$  chain haptoglobin,  $\beta$  chain haptoglobin and Complement Factor H). Analysis by Western blot confirmed the hybridoma screening results and showed that the levels of these three proteins were increased in the plasma of lung cancer patients in comparison to normal controls.

### **Antibody Characterization**

A variety of assays were developed to characterize these monoclonal antibodies, including immunoprecipitation, Western Blot, sandwich ELISA and SPR. ELISA data demonstrated that the three monoclonal antibodies (( $\alpha$  chain haptoglobin,  $\beta$  chain haptoglobin and Complement Factor H) recognized proteins which discriminated lung cancer from normal controls. SPR revealed that the affinity of the monoclonal antibody for haptoglobin was  $200 \times$  greater for the native tetramer (non-reduced) in comparison to the monomer (reduced) form. This demonstrated that the monoclonal antibody recognized the native form of the protein in plasma. Finally the epitope of the anti- $\beta$  chain haptoglobin was to a carbohydrate structure of the protein. Treatment of the haptoglobin with PNGase F to remove the glycans abolished the antigen-antibody interaction.

### **Next Steps:**

Future studies will continue the characterization of the remaining mAb clones produced in this study. Even though identification of the antigens for three of the cancer specific monoclonal antibodies characterized in this study corresponded to high abundant proteins, the success of the initial study suggests that this method could be a powerful approach for rapid generation of disease specific monoclonal antibodies. It is plausible that the levels of abundant proteins reduced using immunodepletion and multi-lectin columns were not sufficient to extend the dynamic range of plasma proteins. Future aims will identify more sensitive protein separation methods which employ targeted fractionation strategies. This could decrease the complexity of the plasma and produce samples containing low level native proteins for immunization and subsequent mAb production. Libraries for mAbs for potential markers to other cancers will also be generated and evaluated.

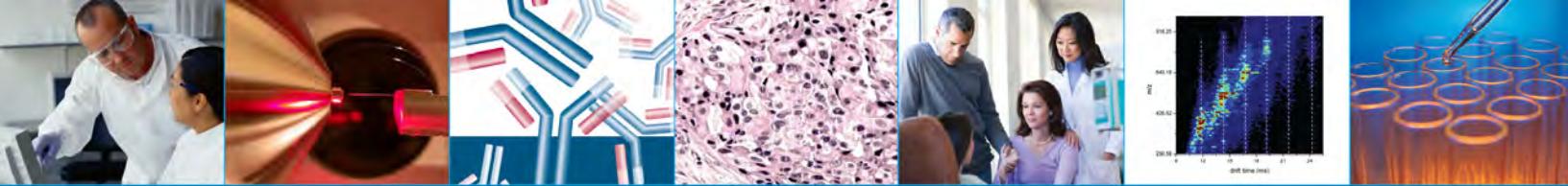
### **Academic/Industry/Research Collaborations:**

Biosystems International (BSI)  
George Mason University  
Northeastern University

### **2.6.3 Developing Synthetic Antibodies for Array-based Cancer Detection**

**Investigator:** John Chaput, Ph.D., Arizona State University

**Project Goals and Significance:**



The ability to monitor the levels of large numbers of proteins for indicators of disease (biomarkers) holds great promise for the detection and treatment of many diseases, including cancer. Although most analytical techniques for detecting protein levels in human serum currently rely on ELISA-based assays or mass spectroscopic analysis, protein capture arrays have gained considerable interest as a future technology for proteome-wide analysis. The current proposal seeks to develop a comprehensive array of high affinity synthetic antibodies (synbodies) that can be used to detect and profile cancer and cancer-related proteins in human blood and saliva. In contrast to immunoglobins, whose high costs and slow production rates limit the availability of suitable quality affinity reagents, the rapid synthesis of artificial antibodies from low cost chemical reagents offers a possible solution to the protein ligand problem. The goal of this proposal is to develop synthetic antibodies to five known cancer biomarkers using a technology that we call SRLL (systematic recombination of ligands and linkers). This technology combines the multiplex capability of SPR with the nanometer-scale precision of DNA self-assembly to identify optimal peptide pairs and peptide pair separation distances needed to produce multivalent binding agents from monovalent ligands previously identified by screening or selection methods. We have termed the ligands that emerge from this process "synbodies".

The specific aims of this proposal are to:

1. Determine the minimum number of selection steps needed to produce modest affinity peptide ligands to five well-established cancer biomarkers.
2. Generate an array of high affinity synthetic antibodies to the five cancer biomarkers targeted in Specific Aim 1.

**Accomplishments:**

Dr.Chaput's R21 grant was recently approved in September 2008. The project is now underway.

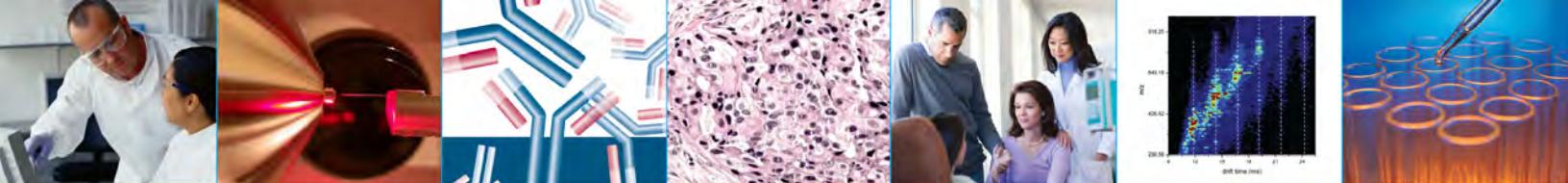
#### 2.6.4 Yeast Single Chain Antibodies as Capture Reagents

**Company: Allele Biotechnology & Pharmaceutical, Inc.  
SBIR**



**Proposal:**

The goal of this project is to develop single-chain variable-fragment antibodies (scFV) against cancer-related proteins. Although monoclonal antibodies are the most widely used reagents for detecting and quantifying proteins, their development is time consuming and expensive. Furthermore, many antigens are non-immunogenic and therefore unusable for antibody generation. The large size of the antibodies may also



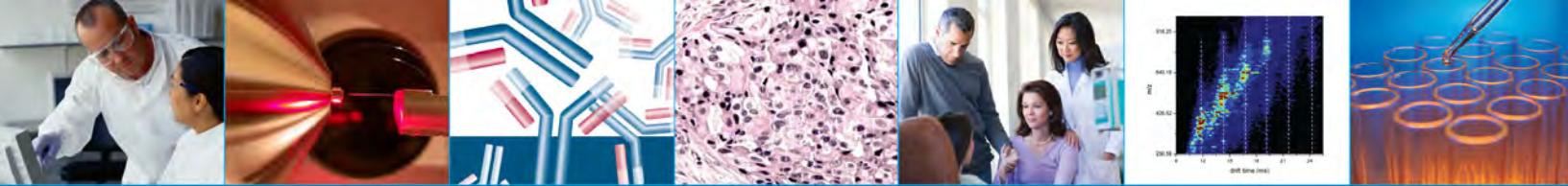
limit their use in cases where more than one antibody competes for closely juxtaposed epitopes. Therefore, scFv can serve as alternative capture reagents with high specificity, high affinity, and small size that can be easily and economically produced. The company proposed to screen (against 10 cancer antigens) a large scFv antibody library composed of genes encoding rationally-designed complimentarily-determining region sequences with designed codons mimicking natural human antibody diversity. The isolated antibodies will be characterized by Western Blot, ELISA and microarray platforms.

#### **Set Milestones:**

- Completion of scFv yeast display library with tailored codons
- Isolation of scFv antibodies for 10 cancer related antigens
  - Enrichment by magnetic beads
  - Selection by flow cytometric sorting
  - Characterization of isolated scFv antibodies – affinity
  - Characterization of isolated scFv antibodies – specificity
- Characterization and evaluation of the isolated antibodies in ELISA assay and/or microarray applications; preparation for larger scale Phase II studies.
  - Identification of scFv antibody pair for ELISA
  - Qualifying ELISA
  - Affinity maturation
  - Application of scFv in microarray

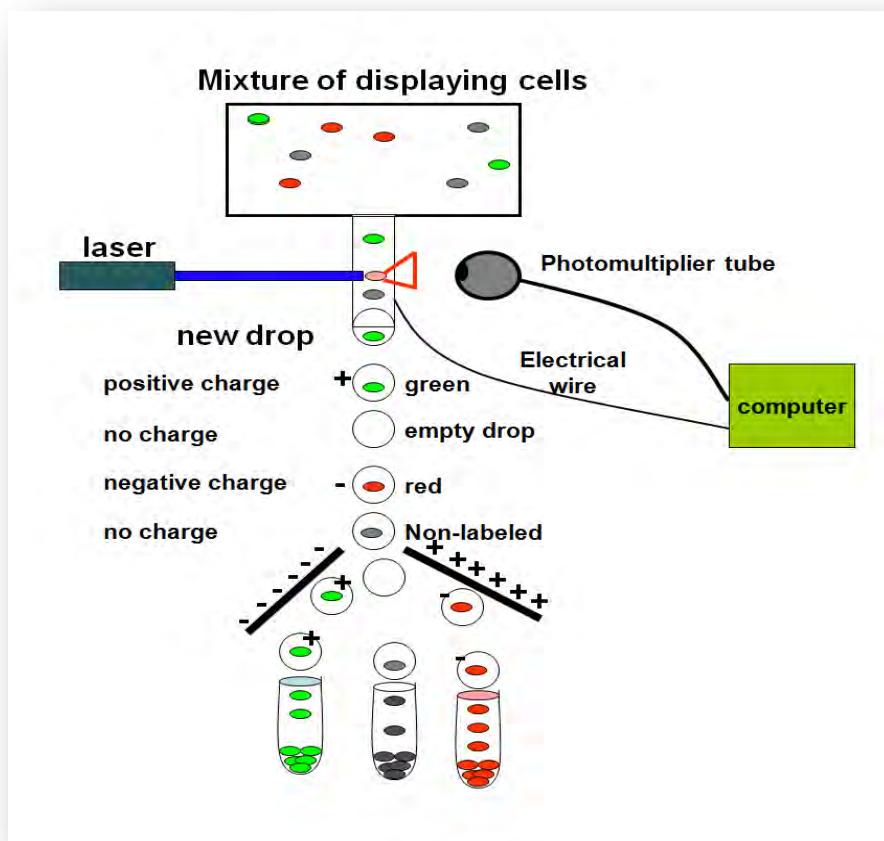
#### **Phase I Results:**

Contractors have isolated multiple high affinity and specificity antibodies against 8 of the 10 proposed antigens, and pools of antigen-specific scFv clones against the other 2. Contractors have also conducted experiments to address antibody pair selection and characterization of isolated antibodies including specificity, Kd, and Western blot.



**Figure 42. Allele FACS flowchart**

Cells displaying different antibody fragments are labeled through binding to fluorescently labeled antigens. Different colors correspond to different binding specificities. The fluorescent signals are converted by the cell sorter to electronic charges that separate cells into collection tubes.

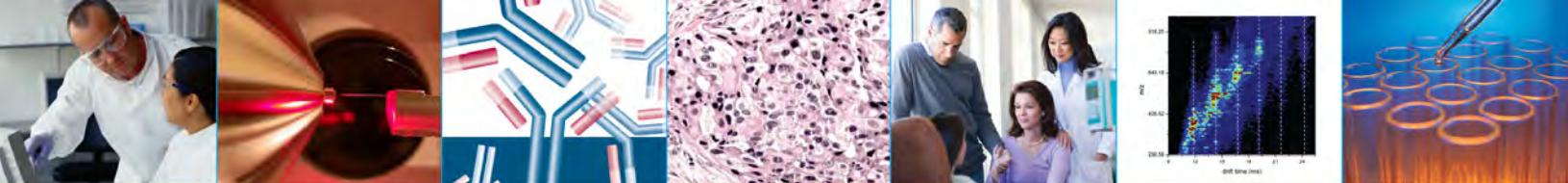


#### 2.6.5 Analysis and Statistical Validation of Proteomic Datasets

**Investigator:** Alexey Nesvizhskii, Ph.D., University of Michigan

##### **Project Goals and Significance:**

Dr. Nesvizhskii is developing computational methods which take advantage of new MS instrument capabilities and lead to increased accuracy and sensitivity. This is particularly important in the identification of low abundance proteins, where the confidence in the identification can be increased through the combination of multiple search tools and through integration of complementary spectral information from multistage MS.



## Accomplishments:

### **Improving Peptide Validation (*sensitivity and robustness*)**

Developed a new class of statistical models that apply the Peptide Prophet method in a variety of settings, with any search tool and without the need for optimization. The models have been implemented into the Trans-Proteomic Pipeline (TPP), which is utilized by the proteomics research community.

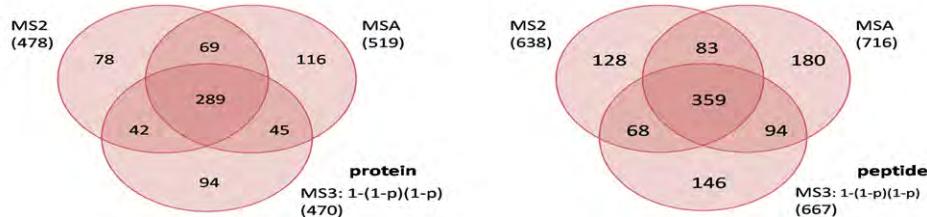
### **Novel Multiple Database Search Tools: Improved Sensitivity**

The use of one data analysis search program can only identify a fraction of the spectra. Variations in algorithms provide different identification results. To leverage this variation, a probabilistic framework tool was developed for combining results from multiple search engines. A significant gain in the number of peptides and proteins (>30% improvement) identified with high confidence was achieved. The approach has led to a computational tool Scaffold (Proteome Software). The model will be available to the research community as part of the iProphet tool currently in development.

### **Building Confidence in Phosphopeptide Identification**

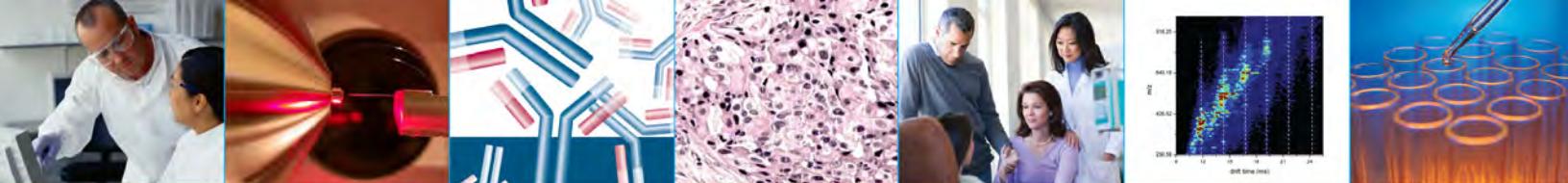
Through the combination of multiple stages of mass spectrometry, such as MS2 (MS/MS) and MS3 data, improvement was observed in the confidence for peptide and protein identification. An advantage to this method is the use of a protein mixture dataset in conjunction with a complex phosphopeptide enriched dataset. Analysis of multistage activation (MSA) revealed that this method is superior to MS2 and MS3 alone and that the combination of different strategies is necessary for optimal results (Figure 43).

**Figure 43. Analysis of Multistage Activation (MSA)**



### **Sequence Tag-Generation: Improved methods**

The use of improved sequence tag generation methods provides quick and sensitive identification of novel and post-translationally modified peptides. This approach provides a >25% improvement for identification of phosphopeptides compared to conventional methods.



### ***Analysis of Label Free Spectral Count Data***

The Q-Spec statistical model for detection of differentially expressed proteins using label free spectral counting method has demonstrated superior performance in comparison to existing methods developed for gene expression analysis. Q-Spec allows investigators to determine the number of replicates required to detect a certain fold change in differential expression. This will be useful for designing label free experiments which utilize spectral counting techniques.

#### **Next Steps:**

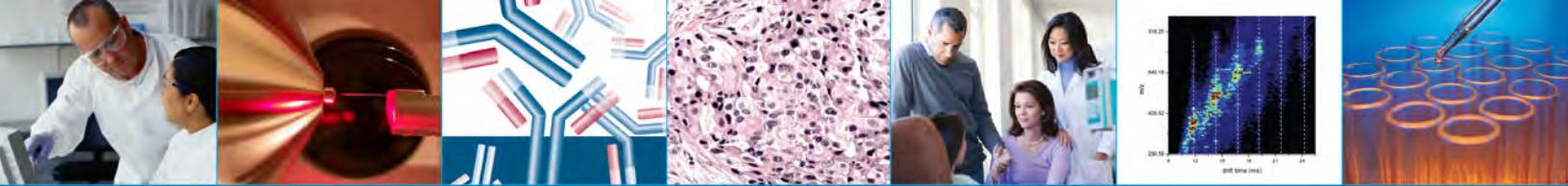
Future studies will focus on the analysis of protein level error rates in large proteomic databases and datasets. The development and implementation of iProphet improves the current method (PeptideProphet/ProteinProphet). Preliminary results demonstrated improved error rate analysis. The laboratory continues to improve normalization schemes through the incorporation of peptides' physico-chemical properties in the analysis. The model will be extended and allow for quantification of differences in protein expression of phosphorylated and glycosylated proteins.

#### **Academic/Industry/Research Collaborations:**

University of Michigan  
Institute for Systems Biology  
Swiss Institute of Technology  
Proteome Sciences Company  
National Institutes of Standards and Technology  
Mount Sinai Hospital  
University of Toronto  
Stanford University  
Boston University

#### **2.6.6 Standard Reference Materials**

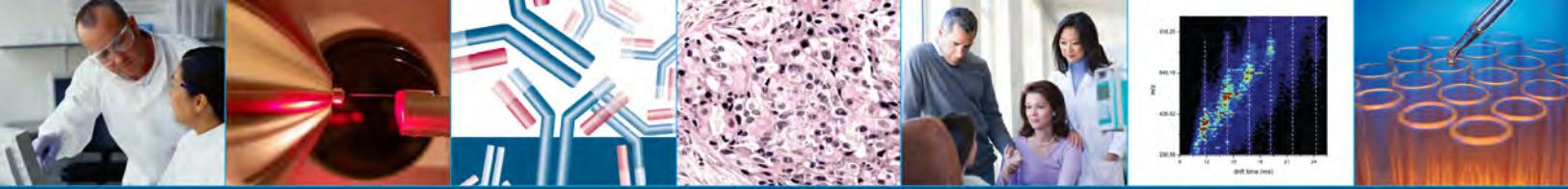
In the first two years of the CPTC program, NIST supported the characterization of proteins and peptides as measurement assessment materials toward the goal of developing a standard reference material to support proteomic analysis by MS and other analytical methods. Having successfully completed the initial experiment involving a 20 protein mixture (NCI-20), NIST is currently preparing materials which more closely simulate biological specimens (tissues/plasma). The first reference material created, the NCI-20, contains human proteins added at prescribed concentration levels. This is unique in that most, if not all, similar commercially available reference materials have proteins which are equimolar. This material was evaluated as part of several inter-laboratory studies within the CPTC program. Inter-laboratory studies have been shown to be an excellent means of identifying potential reference materials which may be of value to the proteomics community.



As a result of the CPTC program, two new products have been approved for SRM development: Yeast Lysate for Proteomic Research (SRM 3953) and Peptide Performance Mixture for MS (SRM 3592) (Table 3). The Yeast Lysate was used in several of the CPTC inter-laboratory studies. The Peptide Performance Mixture will be an aqueous mixture of ~400 synthetic peptides, whose concentration will span approximately 3 orders of magnitude. The Peptide Performance Mixture is being generated and will be used in future CPTC inter-laboratory studies. The generation of SRMs requires stability testing, quality control evaluation and extensive characterization before it is released to the scientific community. In the interim these materials will be released, concurrent with CPTC inter-laboratory study publications and will be available to the scientific community as a SRM-grade material until the official SRM is released.

**Table 3. Two New Products Approved for SRM Development**

Standard Reference Material	Purpose
<b>Yeast Lysate for Proteomic Research (SRM 3953)</b>	Model proteome of moderate complexity for evaluating measurement quality of proteomic evaluations  Tool for the development of new measurement techniques
<b>Peptide Performance Mixture for MS (SRM 3592)</b>	Complex mixture of peptides for evaluating the performance of shotgun proteomics MS instruments coupled to liquid chromatography



## 2.6.7 A Platform for Pattern-based Proteomic Biomarker Discovery

**Investigator:** Denkanikota Mani, Ph.D., Massachusetts Institute of Technology

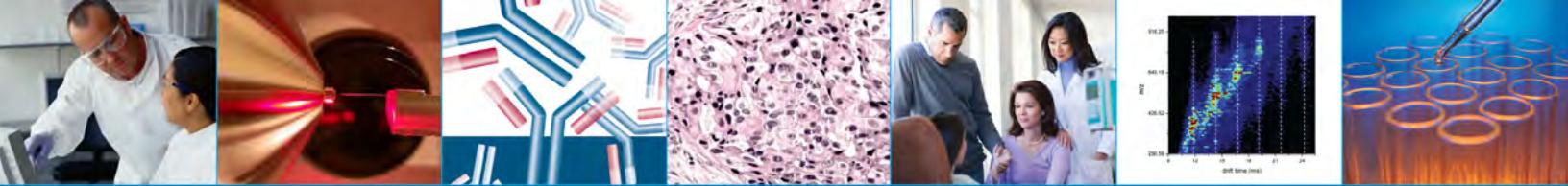
### Project Goals and Significance:

To construct and validate a software system for protein/peptide pattern discovery, this research team will combine peptide identity and pattern information obtained from high resolution and high mass accuracy spectra. Application involves the use of peptide identifications via tandem MS throughout the processing of the data set, while still allowing quantification and comparison of unidentified peptide signals.

### Accomplishments:

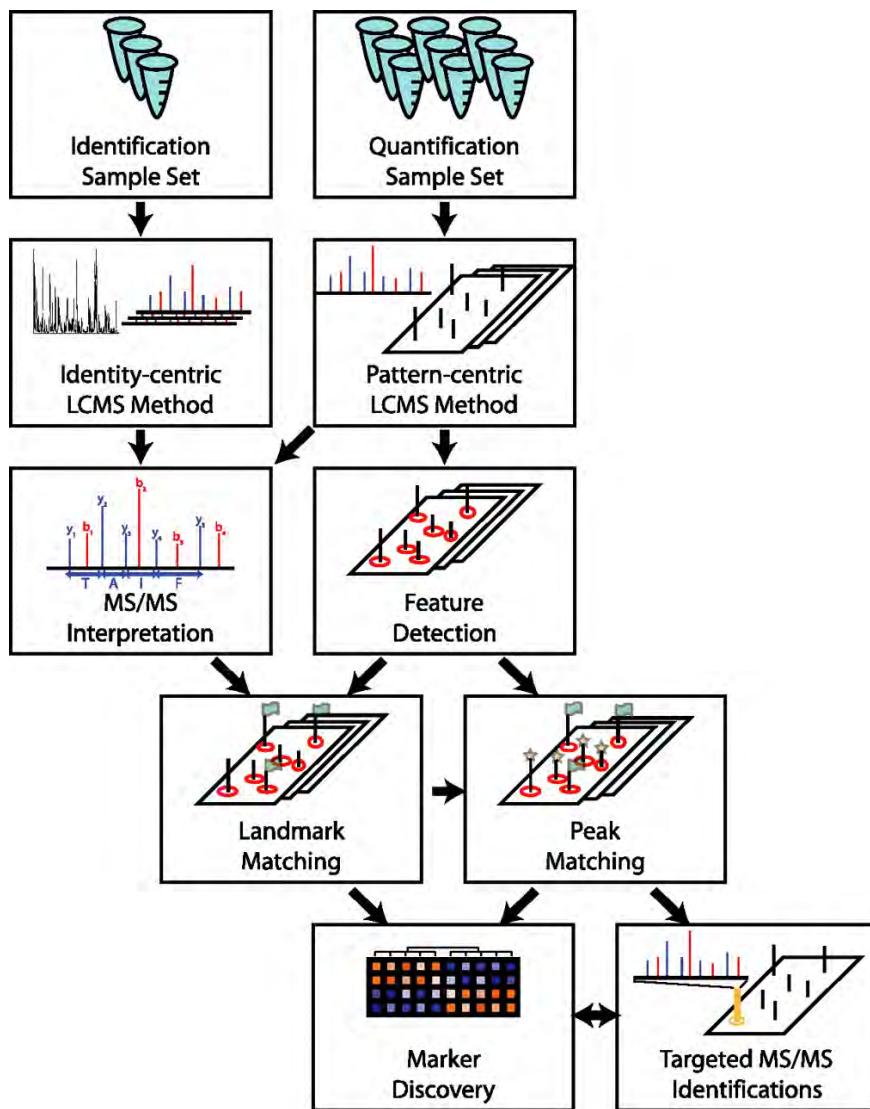
**Implementation of PEPPeR** Quantitative proteomics using LC-MS holds considerable promise for elucidation of basic biology and for discovery of clinical biomarkers. However, effectively harnessing the wealth of information in LC-MS data necessitates going beyond the traditional approach of analyzing only identified peptides. To address these challenges and facilitate robust and high throughput biomarker discovery, Dr. Mani has developed PEPPeR—A Platform for Experimental Proteomic Pattern Recognition. PEPPeR uses high resolution and high mass accuracy LC-MS data from state-of-the-art mass spectrometers, and appropriately combines pattern-based (unidentified peptide peaks) and identity-based (peptides sequenced via MS/MS) information to generate peptide quantitation—thereby extending biomarker discovery to *all* charge identified MS1 peaks. From a computational perspective, the uniqueness of this approach arises from the use of: (i) identified peptides to guide alignment of unidentified peaks; (ii) matching unidentified peaks across multiple samples using mixture model based peak matching; and (iii) adaptive matching tolerances automatically calculated for each experiment.

**Development of fPEPPeR** PEPPeR, in its original form, and all other related algorithms are currently only capable of analyzing one-dimensional LC-MS/MS data. However, it is well known that two dimensional (2D) separations employing orthogonal separation modalities (e.g., reversed-phase (RP) separation of fractions produced by SCX of peptides) greatly increase the number of peptides detected in a sample. A recently-developed extension of PEPPeR, named fPEPPeR, incorporates the very first methodology for processing and computationally reassembling peptide fractions from multidimensional fractionation to facilitate data analysis at the sample level. The method works well despite imprecision of fraction boundaries or other variations during fractionation. In fPEPPeR, the PEPPeR peak-matching algorithm has been adapted to identify the same peptide species (peak) not only across multiple samples, but also across different fractions. fPEPPeR outputs intensity measurements for a common set of peaks spanning all the fractions under consideration. A sample is then computationally reassembled (i.e., defractionated) by summing the intensity measurements for each matched peak across all fractions from that sample. The de-fractionated samples can then be subject to biomarker discovery, class prediction, clustering and other pattern recognition algorithms without regard to the fractionation or any variations therein. This software is freely available as a GenePattern module (<http://www.broad.mit.edu/cancer/software/genepattern/desc/proteomics.html>)



**Figure 44. PEPPeR schematic**

PEPPeR schematic taken from Molecular & Cellular Proteomics 5:1927-1941, 2006.

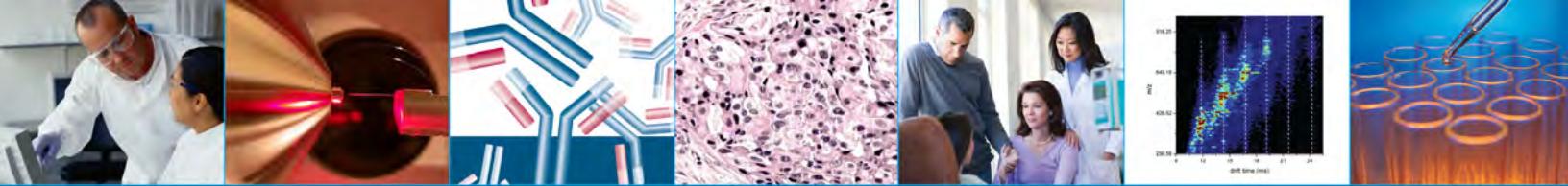


#### Next Steps:

Future work will include continued development and deployment of the PEPPeR and fPEPPeR platforms, with an emphasis on applying the platform to real-world biomarker discovery efforts. Dr. Mani will address challenges beyond biomarker discovery, by addressing computational challenges in biomarker verification and validation.

#### Academic/Industry/Research Collaborations:

Institute for Systems Biology, Seattle, WA



Fred Hutchinson Cancer Center, Seattle, WA  
The Canary Foundation, San Jose, CA.  
Entertainment Industry Foundation  
Gates Grand Challenge Project as carried out at the Broad Institute of MIT and Harvard, Cambridge. MA.

### 2.6.8 PICquant-An Integrated Platform for Biomarker Discovery

**Investigator:** Dennis Templeton, University of Virginia

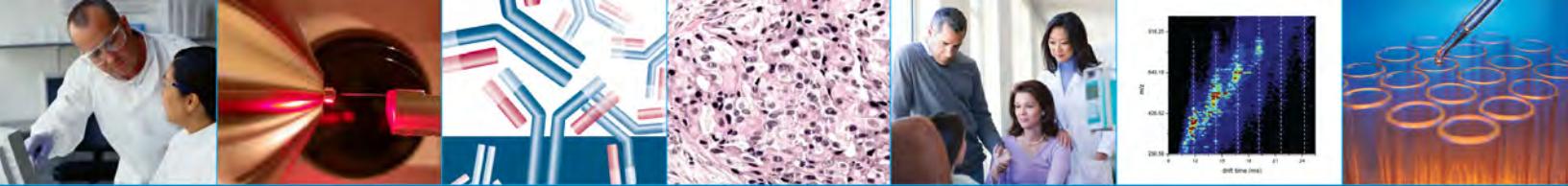
**Project Goals and Significance:**

Dr. Templeton seeks to develop a  $^{13}\text{C}$  phenylisocyanate (PIC) label and a PICquant software platform for accurate identification of peptide peaks in treated urine samples, and eventually, the identification of novel biomarkers for diseases. Urine is an easily accessed specimen that has not been well exploited for cancer biomarker studies. Dr. Templeton has developed a new stable isotope mass tag ( $^{13}\text{C}$  Phenylisocyanate, PIC) that has several advantages for protein quantification in complex mixtures. This platform includes: 1) custom designed software (PICquant) to automatically quantify labeled peaks, 2) a spectrum-comparison algorithm that groups spectra into a Registry of spectra representing unique peptide families, and 3) enhanced peptide sequencing by distinguishing b- and y-ion series in CID spectra. Completing the platform is a clinical registry that links acquired specimens to current and prospective clinical information including outcomes. The registry enables multivariate clustering of disease states with quantified protein families. The mature PICquant platform will provide nearly completely automated data analysis, allowing assembly of numerous patient samples into complete protein abundance profiles akin to gene expression array data. Additionally, the expanding Registry database will be of use to other investigators processing either PIC-labeled or unlabeled peptide spectra.

**Accomplishments:**

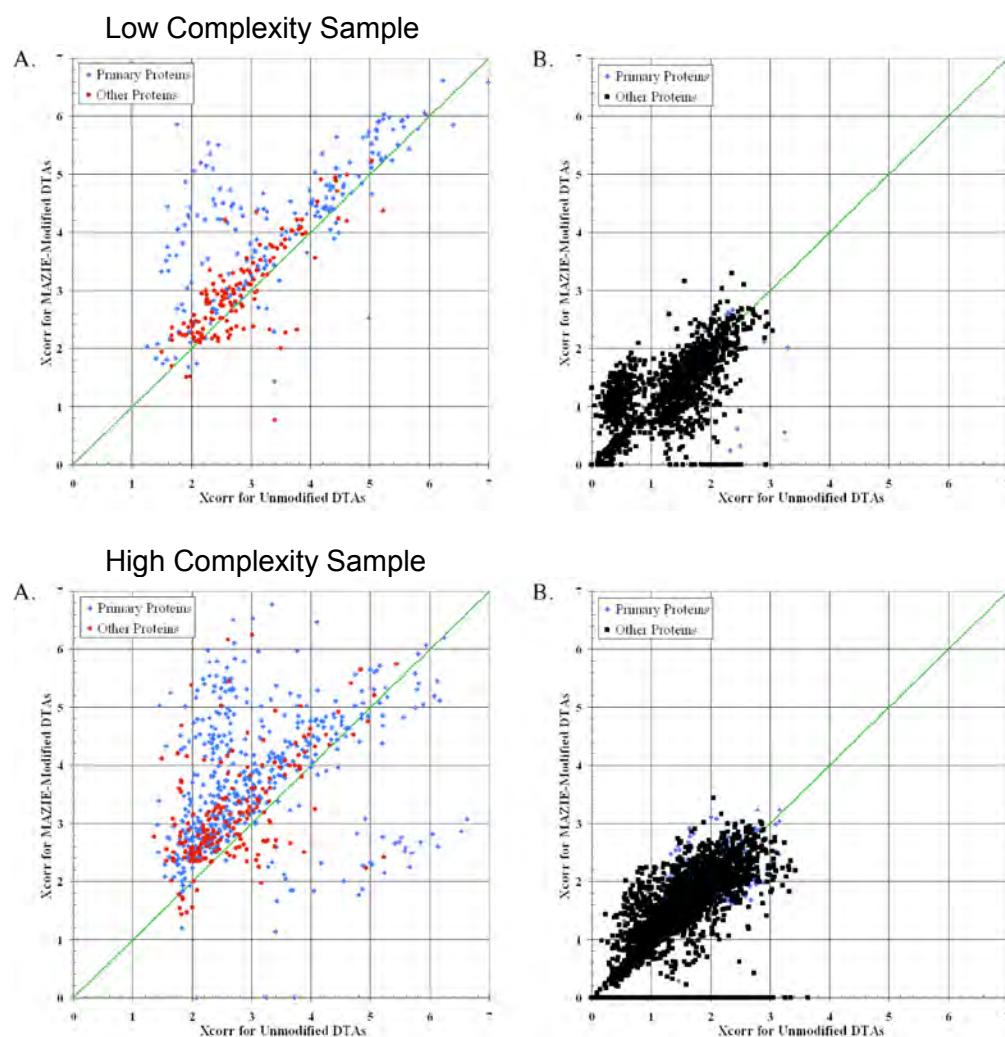
**Developed MAZIE software**

This software improved identification of peptide ion mass and charge, based on the isotopic distribution of peptide ion envelopes. This software will be distributed freely to the research community upon publication. MAZIE was written to integrate into PICquant, because the improved mass and charge information enhances the fidelity of ion clustering. MAZIE is however suitable for general use to enhance database searching for mid-resolution mass spectrometers (Figure 45).



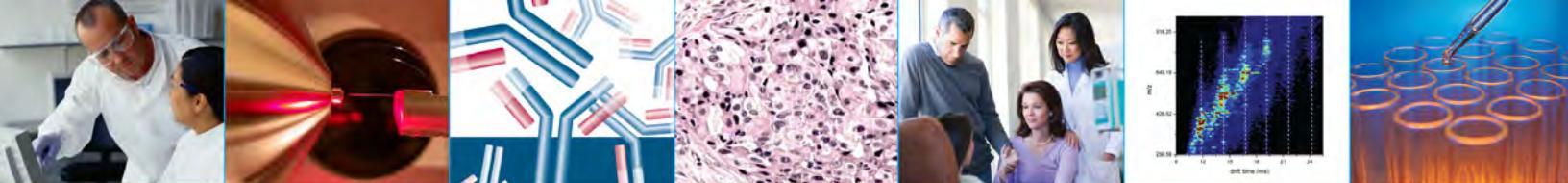
### Figure 45. MAZIE, the Mass and Charge (Z) Inference Engine

MAZIE, the Mass and Charge (Z) Inference Engine, allows superior identification of MS2 spectra using database search engines, in this case Sequest. Proteins from low-complexity gel band samples (top) or highly complex unfractionated urine samples, bottom, were analyzed using MAZIE and Sequest (vertical axis) or Sequest alone (horizontal axis). XCorrs falling above the diagonal line represent scores that are improved using MAZIE. Proteins with XCorr scores above the 5% false positive rate (A) were generally improved, while those below the 5% FPR (B) did not show improvement. Colored spots represent proteins known to be in each sample, or other proteins as indicated.



### Updated PICquant software

The clustering algorithm of the PICquant software has been updated, and this has improved the reliability with which separate measurements of peptide concentrations are



made. Different data acquisition strategies have been compared and one was chosen, in which a pilot case-control sample is fractionated and analyzed in depth, requiring 15 to 20 separate analyses. From these, and integration of data into the PICquant registry, 250 or so 'pathfinder' ions are anticipated to be identified that tentatively show correlation with disease state. Subsequently, individual cases and controls will be quantified using acquisition focused on these specific pathfinder ions.

#### **Developed a new reagent**

A new reagent (tentatively called TissueSolv) was developed that is capable of reversing many of the protein-protein crosslinks in formalin-fixed tissue. This reagent is superior to previous methods that mainly solubilize the proteins, because it results in peptides whose masses match those predicted by protein databases. This approach enables use of archival samples for validation of potential marker expression by MS.

#### **Next Steps:**

Acquisition of clinical samples from women, obtained prior to biopsy-proven breast cancer has been completed. A preliminary 'catalog' of potential low- to mid-probability cancer markers is being completed. From this list, a set of mass analytes will be generated to scrutinize 40 case samples and 40 controls from cancer affected urine. Based on pilot analyses, this new strategy of identifying candidate peptides in pooled cases and controls first (about 250), then quantifying these specifically in the non-pooled cases and controls using instrument 'inclusion lists', will allow for a much higher chance of accurately quantifying peptides of potential interest than when using data-dependent, stochastic acquisition of non-fractionated samples. Initially the MS analysis of samples will occur followed by statistical analysis.

#### **Academic/Industry/Research Collaborations:**

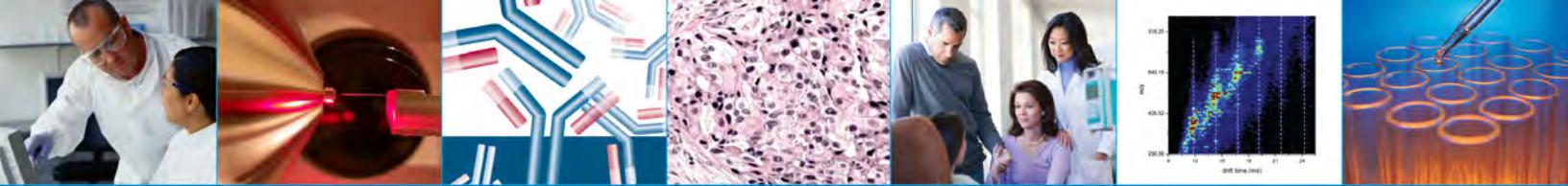
University of Virginia

### **2.6.9 New Proteomic Algorithms to Identify Mutant or Modified Proteins**

**Investigator:** David Tabb, Ph.D., Vanderbilt University

#### **Project Goals and Significance:**

Dr. Tabb is producing database search identification tools that augment proteome informatics in several key ways. The tools include improvement to peptide identification scoring, improving the discrimination of good versus bad matches, thus increasing sensitivity and specificity of proteome identification. These tools also assess error rates, making false discovery rate evaluation central to the process of data mining in proteomics and which contributes to the stability of data reporting. Most proteomic laboratories will over report protein counts due to shared peptides. An open sourced data mining tool was produced which resolves this issue. A key advance has been the development of DirecTag, a tag-based identification which has been shown to be an accurate way to recover sequences from tandem mass spectra.



## Accomplishments:

### DirecTag: Novel tool for sequence tag inference

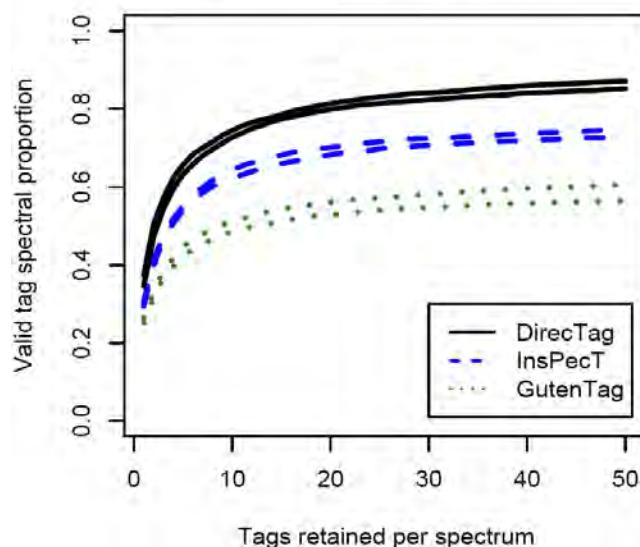
Identification of peptides through sequence tagging using automated sequence tag inference. The algorithm has been evaluated on a diversity of MS instruments, from TOF/TOF to quadrupole ion trap. All results indicate that this algorithm generated correct tags for a higher proportion of identifiable spectra in comparison to other methods tested (Figure 46). The degree of accuracy is important to ensure that these tags can serve as a basis for identification of the remainder of each spectrum. DirecTag has been released with source code to the research community.

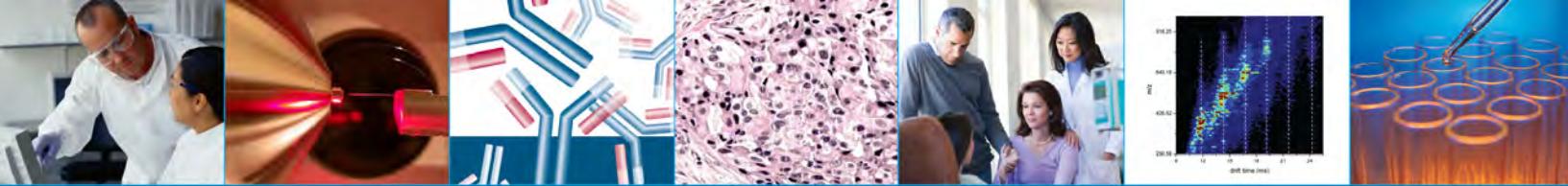
The real value of sequence tagging in cancer is the identification of mutations in cancer samples that have been missed by database search. The process of tag reconciliation can allow amino acid changes to either side of the inferred sequence. In this process the tag sequences for an MS/MS are reconciled against the protein sequences from the database. The “TagRecon” software conducts this process using the same scoring algorithm as in MyriMatch. Preliminary results indicate that TagRecon complements standard database searching powerfully, even when mutations and modifications are ignored. Therefore combining TagRecon and MyriMatch search results increases confident peptide identification. The two identification approaches show both commonality and complementarity. Validation of the mutation detection capabilities in cancer related datasets is under investigation.

### Figure 46. DirecTag generates more accurate tags than InsPecT or GutenTag

*In these eight Orbitrap RPLC analyses of yeast lysate, DirecTag consistently generates sequence tags that agree best with the Sequest identifications of the spectra. This difference is consistent when fewer than 10 tags are retained per spectrum and when more than 30 tags are retained per spectrum. When 20 tags are retained per spectrum, DirecTag generates a correct tag for approximately 80% of identifiable spectra.*

Sequest-Yeast: high/low





### ***MyriMatch***

MyriMatch was developed as a powerful match scoring technique and since its introduction it has been utilized in two other algorithms developed at Vanderbilt and which are under consideration for other tools in other institutions. MyriMatch makes more effective use of fragment ion intensity in comparison to X!Tandem Expect and Sequest XCorr and is robust against noise peaks. MyriMatch has been selected as the standard search engine for processing the data sets of the CPTAC Unbiased Working group.

### ***IDPicker***

The IDPicker tool enables users to organize experimental data into complex hierarchies. It was developed for protein assembly and has proven to be invaluable in generating tables of spectral counts that can be used for identifying candidate biomarkers in large cancer data sets. It has also been instrumental in organizing the complex datasets from the Unbiased Discovery inter-laboratory studies.

### **Next Steps:**

Future studies will include the submission of several manuscripts to validate and demonstrate these new technologies (ScanSifter, TagRecon and IDPicker algorithms). Many file formats are employed in proteomics. ScanSifter enables the translation of mass spectrometry data among these formats in a user friendly graphical user interface. The “sifter” is its ability to screen out spectra that are unlikely to be identified and remove redundant duplicate spectra from a collection. DirecTag generates high quality sequence tags that enable the identification of mutated and modified peptides from cancer samples. This is accomplished through the use of TagRecon where tools to recognize the tags against the sequence database are utilized. Further modifications to IDPicker are being employed and validated to enhance its utility.

### **Academic/Industry/Research Collaborations:**

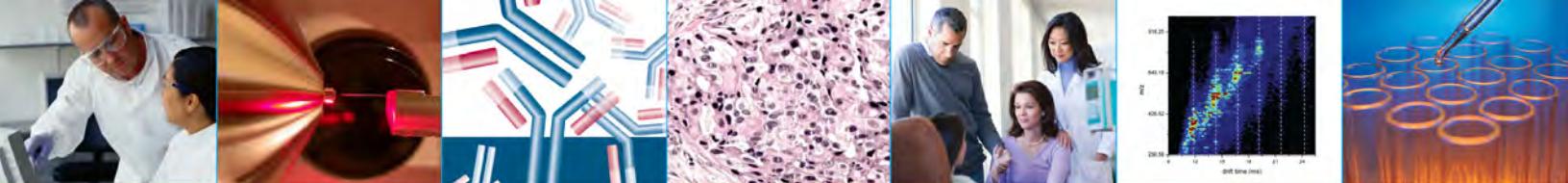
University of Michigan  
Institute for Systems Biology  
National Institutes of Standards and Technology  
Association for Biomedical Resource Facilities  
Buck Institute  
Kansas State University

## **2.6.10 Enhancement of MS Signal Processing Toward Improved Cancer Biomarker Discovery**

**Investigator:** Dariya Malyarenko, Ph.D., College of William and Mary

### **Project Goals and Significance:**

To increase the effectiveness of cancer protein/peptide detection from label-free Matrix Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) mass spectra for



verification and identification, Dr. Malyarenko is developing novel computational tools that can be used across all laboratories employing this MS technology.

Her team's long-term goal is to deploy an optimized data acquisition protocols and enhanced signal processing tools for the improvement of sample preparation to streamline broad range MS mining of proteomes (e.g., through MS imaging and for protein digests) for functionally important molecules (or their modifications) related to different types of cancer. These computational methodologies will be modified to be transferable to a range of high-throughput TOF technologies appropriate for the studies of population effects for verification and further clinical trials.

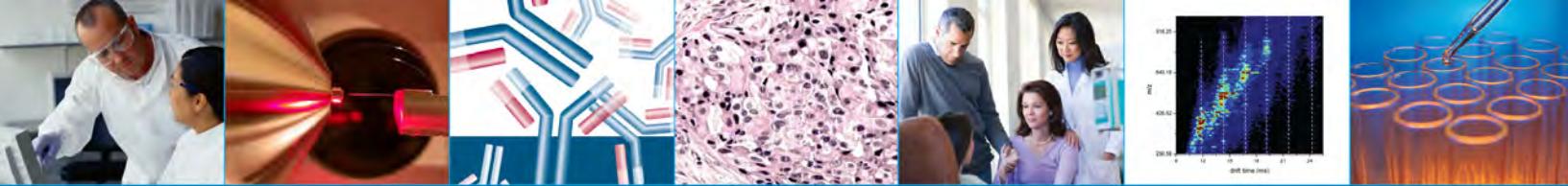
#### **Accomplishments:**

**Developed new models for baseline, instrumental noise, and peak shape for MALDI-TOF mass spectra** Using mass spectra of pooled serum and calibration protein mixtures collected alongside clinical samples, Dr. Malyarenko developed new peak-fitting models for three operating modes of the Bruker Ultraflex MALDI-TOF MS.

**Optimized acquisition parameters** Her team developed a procedure to systematically record instrument parameters in order to minimize variability and bias in future experiments. Signal processing parameter optimization was performed by minimizing uncertainty estimates for signal locations. Each signal processing step included model parameters determined to be appropriate for the particular instrumental acquisition settings used. Optimizing these parameters enhanced the instrument's sensitivity.

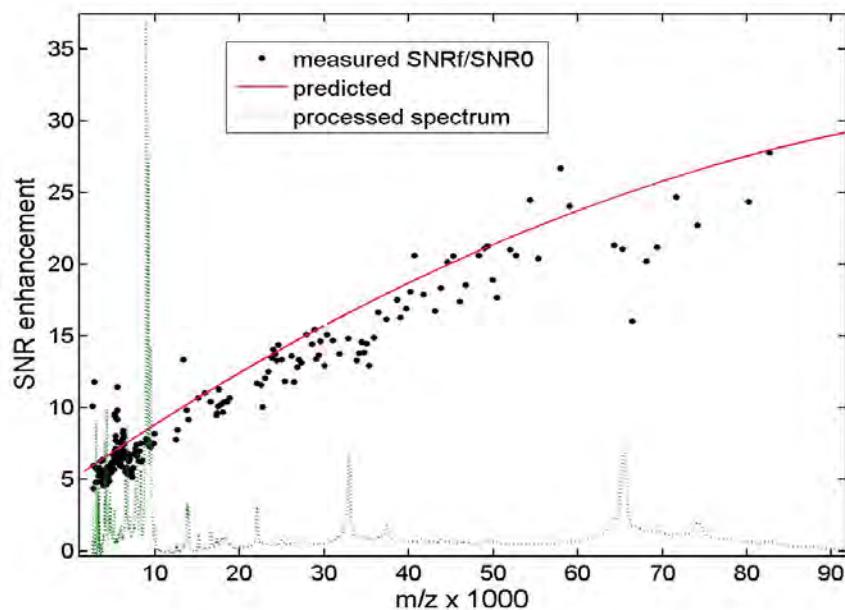
**Developed signal processing libraries** Using the statistical analysis programming language, R, Dr. Malyarenko's team combined several characteristics of mass spectra to create improved signal processing libraries. These enhanced libraries provided tenfold increases in data compression speed and sensitivity. This improved sensitivity of signal detection, especially over the broad mass range, effectively extended the range of useful data for each spectrum.

**Developed analytical framework for detection of multiple states of molecular ions** Analytical framework and computational algorithms were developed for automated detection of multiply-charged and singly-charged multimer states of molecular ions in the TOF domain. These tools enabled a fully automated self-calibration procedure, which immediately achieved *m/z* assignments with 10-fold accuracy in protein standard and pooled serum. The achieved calibration accuracy was adequate for aligning calibration standard spectra collected in three mass ranges. Deconvolution of ionization state intensities into molecular ions enhanced signals in the molecular spectra, decreased correlation among signals across serum spectra, and reduced the dimension for variable selection by more than 3-fold.



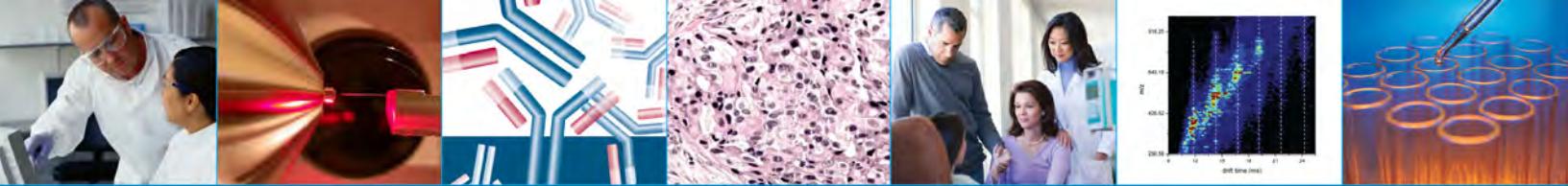
**Figure 47. Measured (points) and predicted (line) sensitivity enhancement for a processed linear TOF-MS spectrum of pooled serum collected using the Bruker Ultraflex III instrument**

The enhancement is measured by taking the ratio of the signal-to-noise ratio for peaks detected in the processed spectrum above SNR=10 (SNR<sub>f</sub>) to the SNR<sub>0</sub> of corresponding signals in the unprocessed data. The processed spectrum (dotted green) is overlaid to guide the eye. The SNR enhancement is more than 10-fold for signals above 10 kDa, where the combination of constant and quadratic integrative down-sampling followed by optimal linear filtering is applied. A smaller enhancement (by a factor of ~7, on average) is observed for signals below 10 kDa, where no quadratic down-sampling (just constant IDS and OLF) was done.



#### Next Steps:

Future work will include the development of tools for automated discovery of diagnostic proteins. The work plan includes characterization of amplitude statistics for data normalization, automated discovery of diagnostic variables using classification, complete workflow integration and testing of caBIG compliance, and TOF/TOF identification of molecular markers. The goal is to enable automated screening of reconstructed molecular profiles for diagnostic patterns using normalized, contrast enhanced heat maps and statistical significance tests. Toward automated selection of diagnostic proteins, mutual information and Bayesian theory will be applied to investigate dependencies between measured molecular abundances and clinical groups. This probabilistic approach will naturally account for uncertainties in measured intensities within the width of a population distribution and will generalize better for new test samples. To further improve identification, chromatographic protein purification and LC-MS/MS will be employed. Identified proteins will be verified and quantified using standard immuno-techniques.



Ultimately, Dr. Malyarenko's team will continue toward combining the analysis of proteomics data with gene expression to obtain a systems view on the molecular level. On the clinical side we will continue the studies to define key cellular proteins involved in mediating cancer-specific changes and disease sub-typing. This will enable future longitudinal studies that seek to determine the role these discovered biomarkers play in the long-term transformation process.

**Academic/Industry/Research Collaborations:**

Eastern Virginia Medical School

INCOGEN

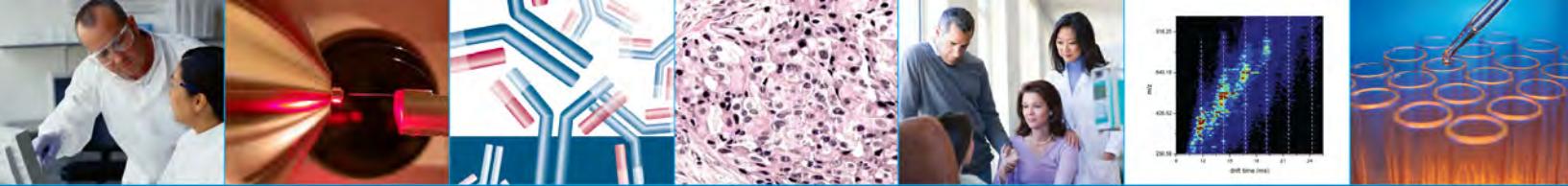
NCI's Early Detection Research Network

#### 2.6.11 Computational Tools for Cancer Proteomics

**Investigator:** Katheryn Resing, Ph.D., University of Colorado at Boulder

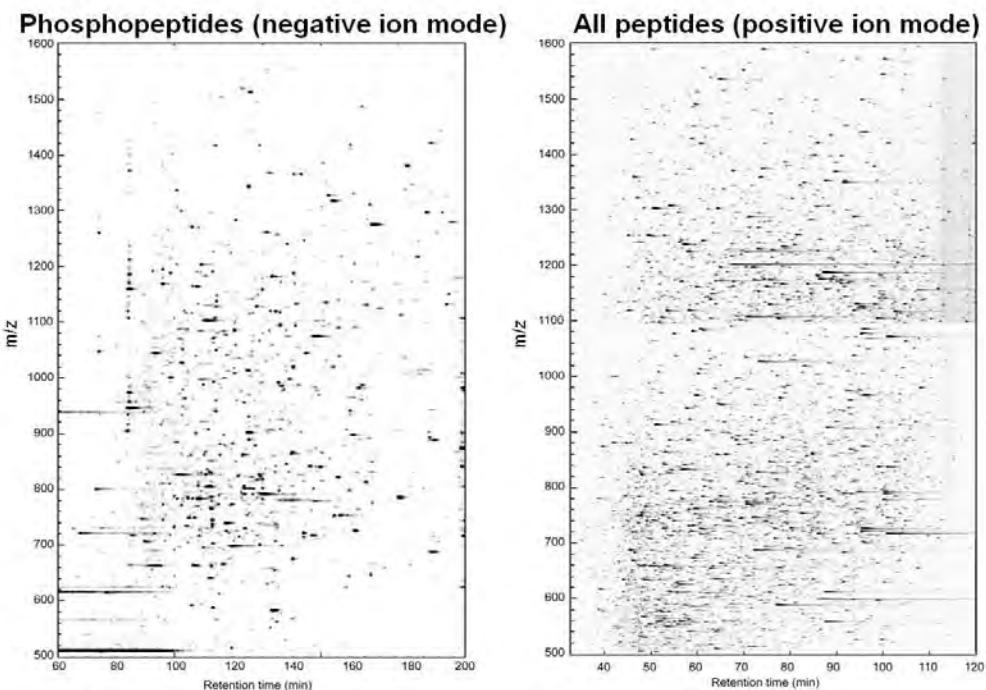
**Project Goals and Significance:**

Dr. Resing is developing new computational methods to profile protein expression and phosphorylation changes in response to signaling pathways and disease states, directly supporting studies of melanoma and prostate cancer. Shotgun proteomics using multidimensional LC/MS/MS approaches that are based on peptide gas phase fragmentation, such as MudPIT, have proven effective in identifying proteins in complex samples. However, serious limitations with respect to depth of sampling proteins in complex mixtures, accuracy of assigning peptide sequences to MS/MS spectra, ambiguities in distinguishing protein isoforms, quantification of protein abundances, and characterization of post-translational modifications, such as phosphorylation exist. In addition, methods are needed to handle problems arising with complex mixtures, such as peaks that overlap in mass and elution, peptides eluting in many fractions during multidimensional separation, and clustering of peptides/proteins based on multivariate measurements. Dr. Resing is developing new computational tools to create an integrated software system which will address these issues.



**Figure 48. A label-free method for phosphoproteomics profiling, using negative ion mode MS**

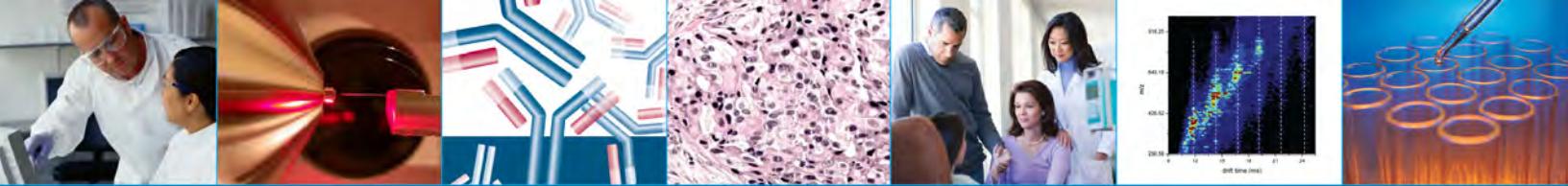
A method developed by Carr et al (1996) scans loss of a phosphopeptide marker ion ( $\text{PO}_3^-$ , 79 Da) in negative mode, and then switches polarity to positive mode to acquire the MS/MS spectrum for sequence identification. This method had not been applied to large scale studies, which we achieved by implementing key improvements to the technology. Left panel: Phosphopeptide candidates identified in negative mode. Right panel: All peptides, detected in positive mode. Comparing the panels shows selectivity of negative -79Da precursor ion scanning to identify phosphopeptide candidates, which were then quantified by label-free methods. In work under review, we applied this technology to screen phosphosite targets of oncogenic B-Raf signaling in melanoma. The method is able to screen for regulated phosphorylation without stable isotope labeling, and represents a promising approach for studies of human specimens.



### Accomplishments:

#### Refinement of Serac software

The suite of software algorithms for quantifying large scale phosphoproteomics datasets which incorporate protein fractionation prior to reversed phase LC/MS/MS analysis with -79 precursor ion scanning has been refined. Feature detection and matching for quantitation of phosphopeptide ions is performed with software designed using the open source C++ library OpenMS (Lange et al., 2007). Once features are detected, a pose clustering approach for matching features across adjacent fractions is utilized. The quantitative phosphopeptide profiles are constructed from the resulting pairwise connection maps using graph traversal algorithms written in the statistical language R.



This method works well for both technical and biological replicates, it correctly aligns > 90% of the peaks.

#### ***Examination of a new method for calculating theoretical MS/MS spectra to facilitate peptide identification***

Two areas of interest: 1) Use of spectra in a spectral library search algorithm and 2) develop an infrastructure for improving modeling. For the first area, a new approach to searching by using a library of previously observed spectra was tested. This provided faster searching than methods such as Mascot or Sequest which search a sequence library. In the second area, studies to improve the Zhang theoretical spectra have been initiated. The use of Dr. Resing's MAE data mining software allows identifications of specific cleavages that were not well predicted. New mechanisms are now being added to the Zhang model to accommodate a new mechanism at this cleavage site, improve predication of the dehydrated/deammoniated ions and develop an alternative approach to determining the charge of the fragment ions.

#### ***Development of statistical and computational methods to support phosphopeptide analysis in complex samples***

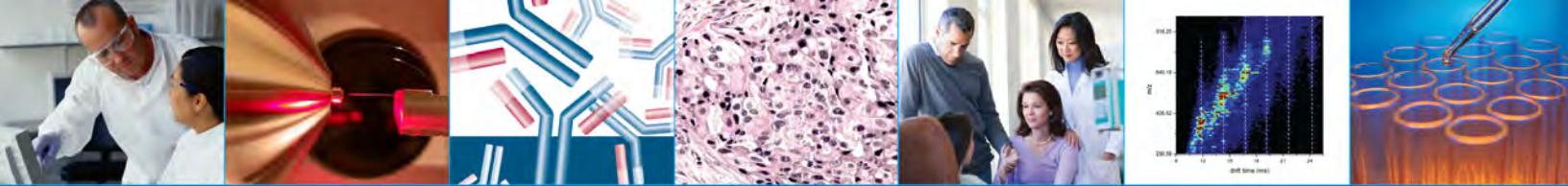
A large phosphopeptide dataset generated from samples separated by Mono Q chromatography was analyzed. Using a workflow based on the OpenMS tools and internal software for global pairwise alignment across the individual LC/MS analyses for each Mono Q fraction. A novel matching metric was also developed which used the similarity of MS/MS between two candidate matched features, even when they have not been identified by the search program, considerably improving the fidelity of the matching in the presence of noise. A third component was an interface to aid in manual validation of results. More than 800 phosphopeptides have been manually validated and qualified in the Mono Q dataset.

#### ***Development of novel label-free strategy for phosphopeptide profiling***

In large scale studies of complex samples, changes in phosphorylation are typically quantified by metabolic or chemically labeling proteins in cells with isotopically distinguishable amino acids. Samples are mixed, phosphopeptides are enriched on various affinity supports, followed by LC/MS/MS. Dr. Resing developed a method for profiling phosphopeptides that does not depend on enrichment, and could be performed quantitatively in a label-free manner. The approach uses a precursor ion scanning method to detect phosphopeptides by their signature fragment ions of  $m/z$  -79, due to loss of  $\text{PO}_3^-$  (Carr et al., Anal. Chem. 1996). Her approach expands on this technology to where phosphoprotein profiling can now be achieved in complex samples without stable isotope labeling, and provides an important new approach for analyzing oncogene targets in human tissue samples.

#### **Next Steps:**

Future studies will evaluate methods for signal processing for the QTrap, to further improve analysis of low intensity ions. Additionally, new datasets from other melanoma cell lines where the WNT signaling pathway, in addition to the B-Raf pathway will be



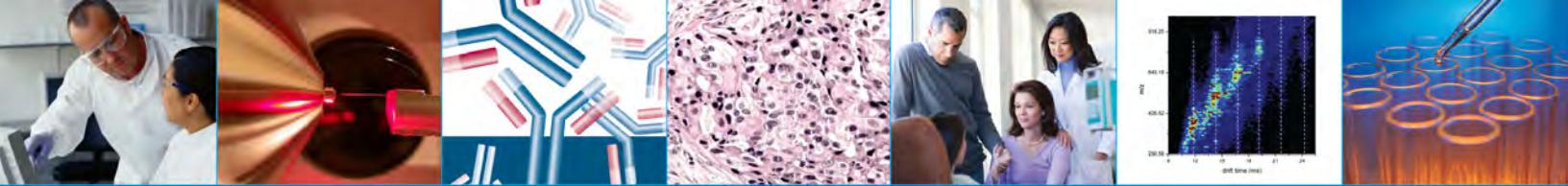
collected and evaluated. Using these datasets, a method for reporting results on phosphorylation and stoichiometry which allows clustering analysis will be developed.

**Academic/Industry/Research Collaborations:**

Amgen Inc.

Arizona State University

University of Colorado Health Sciences Center



### 3. Training Opportunities in Emerging Technologies

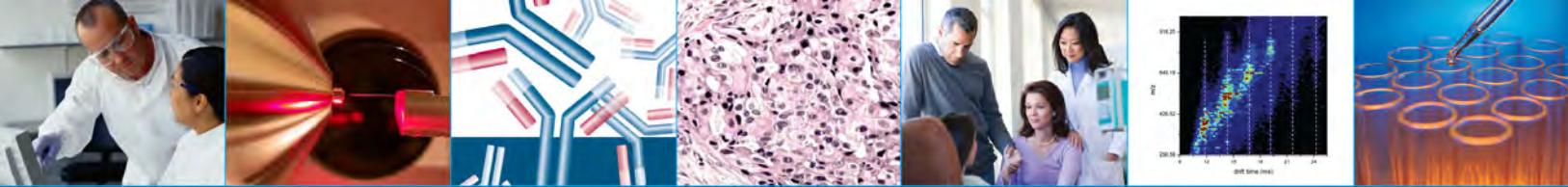
*LeeAnn Bailey, Ph.D.  
Program Director  
Diversity Training Branch  
Center to Reduce Cancer Health Disparities  
National Cancer Institute*

The field of emerging technologies has been identified by the Center to Reduce Cancer Health Disparities (CRCHD) and the CPTC as an increasingly important area for training. It is essential that investigators from racial/ethnic groups and individuals from disadvantaged populations become involved in emerging technology development because of its potential to advance the field of cancer research, and by extension, support the NCI strategic objective to overcome cancer health disparities. By definition, underserved investigators are those who meet one or more of the following criteria: Individuals from racial and ethnic groups that have been shown to be underrepresented in cancer-related biomedical, behavioral, clinical, or social science research; individuals with disabilities; first generation college graduates; individuals who come from a family with an annual income below established low-income thresholds; and individuals who come from a social, cultural, and/or educational environment, such as that found in certain rural or inner-city environments.



The Emerging Technology Continuing Umbrella of Research Experiences (ET CURE) program addresses the need for a diverse cancer research community in the 21st century that reflects the nation's "ethnic heterogeneity", and is sensitive to the significant disparities in cancer health across diverse populations. The overarching goals of the Emerging Technologies Continuing Umbrella of Research Experiences (ET CURE) initiative are to:

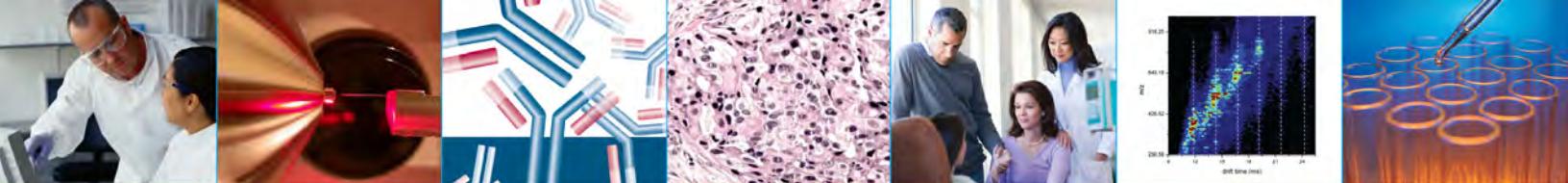
1. Create a pipeline of underserved students and investigators in the fields of emerging and advanced technologies;
2. Increase the number of scientists from underserved populations with training in the elective disciplines of focus, such as nanotechnology;
3. Enhance application of emerging technologies to cancer research through increased training and educational opportunities; and
4. Foster academic, scientific and multi-disciplinary research excellence, culminating in the emergence of a mature investigator capable of securing competitive advanced research project funding.



## 4. Appendix

### 4.1 Organizations Participating in the CPTC Initiative

Accacia International, Inc.  
Allele Biotechnology & Pharmaceuticals  
Argonne National Laboratory  
Battelle Pacific Northwest Laboratories  
The Broad Institute of MIT and Harvard, Proteomic Platform and Cancer Program  
Buck Institute for Age Research  
California Pacific Medical Center  
College of William and Mary  
Developmental Studies Hybridoma Bank at the University of Iowa  
Discovery Park at Purdue University  
Emory University  
Epitome, Inc.  
European Bioinformatics Institute  
Fred Hutchinson Cancer Research Center and its clinical and research partners, the University of Washington and Children's Hospital and Regional Medical Center  
Harvard Institute of Proteomics  
Harvard University and its affiliated hospitals (including Dana-Farber Cancer Institute and Massachusetts General Hospital)  
Hoosier Oncology Group  
Human Protein Atlas (KTH – Royal Institute of Technology; Stockholm, Sweden)  
Indiana University  
Indiana University School of Medicine  
Indiana University – Purdue University Indianapolis  
Institute for Systems Biology  
Lawrence Berkeley National Laboratory  
Massachusetts Institute of Technology  
Memorial Sloan-Kettering Cancer Center  
Meso Scale Diagnostics  
Michigan State University  
Monarch Life Sciences  
National Cancer Institute – Center for Cancer Research Tissue Array Program  
National Cancer Institute – Frederick Advanced Technology Program  
National Institute of Standards and Technology  
New York University Medical Center  
Northeastern University  
Plasma Proteome Institute  
Predictive Physiology and Medicine, Inc.  
Purdue University  
Quadraspec, Inc.  
Rules-Based Medicine, Inc.  
Sequenom, Inc.



## 4.2 Applied Proteomic Platforms and Computational Sciences Investigator Biographies

### Xiaolin Gao, Ph.D., University of Houston

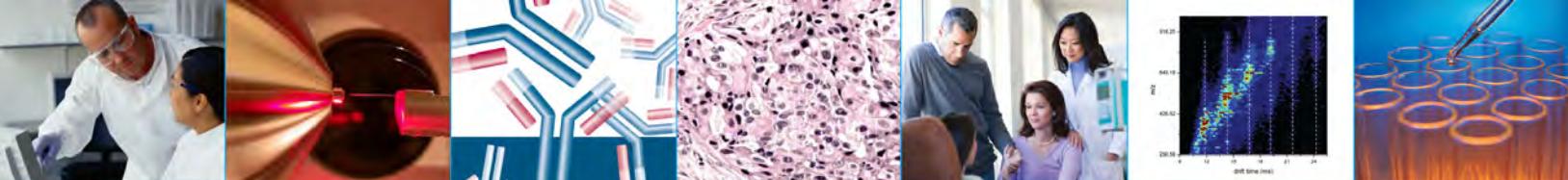
Dr. Gao is a Professor of Biology and Biochemistry and Chemistry, and Director of the Keck/IMD NMR Center at the University of Houston. Pursuing research in the interdisciplinary areas of chemistry and biology, she developed novel methods for miniaturizing massively parallel synthesis of biomolecules on high-density, microfluidic microchips and multiplex biochemical assays on the same chip platform; established the microchip based gene synthesis for many genomic and proteomic applications; and demonstrated microchips to be used as pico-liter titer plates for ultra-high throughput quantitative measurements of binding and enzymatic activities. Dr. Gao's second major research area is solution structures and studies of nucleic acids, proteins, and biomolecular complexes. Dr. Gao holds a B.S. degree from the Beijing Institute of Chemical Engineering and Ph.D. degree in Chemistry from Rutgers University; she did postdoctoral work in NMR-based structure biology at Columbia University Medical School. Before joining the faculty at the University of Houston, she was a Principal Investigator at the Glaxo Research Institute. Dr. Gao is a devoted educator, mentor, and founder of several biotechnology companies.

### Barry L. Karger, Ph.D., Northeastern University

Dr. Karger holds a Ph.D. in chemistry from Cornell University and a B.S. in chemistry from the Massachusetts Institute of Technology. His research focuses on the application of microscale separations to problems of biological significance. Current research involves (1) biomarker discovery and monitoring; (2) ultrasensitive LC-ESI-MS using narrow bore monolithic (20 µm i.d.) and porous layer open tubular (PLOT, 10 µm i.d.) columns coupled to MS; (3) laser capture microdissection analysis of cervical scrapings in conjunction with Pap smear tests; (4) ultratrace characterization of complex proteins, including post-translational modifications using extended range proteomic analysis (ERPA); (5) ultratrace fast MALDI-TOF/TOF MS using a 2 kHz laser; and (6) monoclonal antibodies in biomarker discovery and monitoring. The laboratory provides an interdisciplinary environment for research in collaboration with academic, medical and industrial scientists.

### Joseph A. Loo, Ph.D., University of California Los Angeles

Dr. Loo is a Professor in the Department of Biological Chemistry, David Geffen School of Medicine, and in the Department of Chemistry & Biochemistry at the University of California, Los Angeles (UCLA), and is the Director of UCLA Jonsson Comprehensive Cancer Center Mass Spectrometry and Proteomics Shared Resource. He is also a member of UCLA/DOE Institute for Genomics and Proteomics, the UCLA Molecular Biology Institute, and the UCLA Jonsson Comprehensive Cancer Center.



His research interests include the development of bioanalytical MS methods for the structural characterization of proteins and their application for proteomics and disease biomarkers. He is also interested in development and application of electrospray ionization mass spectrometry for the study of noncovalently-bound macromolecular complexes and their interactions with other binding partners and ligands.

### **Daniel B. Martin, M.D., Institute for Systems Biology**

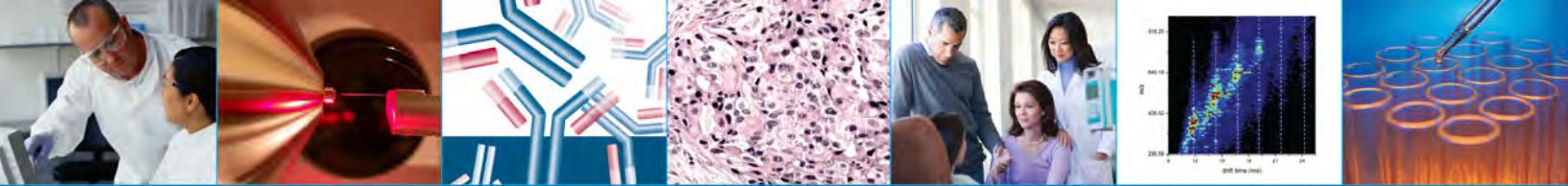
Dr. Dan Martin joined the Institute of Systems Biology as an assistant professor and key leader in its proteomics facility. Dr. Martin is a practicing oncologist at Harborview Medical Center, an associate in Human Biology at the Fred Hutchinson Cancer Research Center and a Lecturer in the University of Washington's Division of Hematology. Dr. Martin, who is a graduate of Yale University Medical School, focuses his research in three areas: The first is to use cutting edge instrumentation to identify new diagnostic markers for prostate cancer. The second is to perform a focused proteomic analysis on the androgen receptor to characterize the proteins participating in this signaling pathway. Finally Dr. Martin has been fascinated by mass spectrometry instrumentation, mechanisms of peptide fragmentation, and computational aspects of mass spectrometry-based proteomics and have pursued a variety of small projects along these lines.

### **Junmin Peng, Ph.D., Emory University**

Dr. Peng is an Assistant Professor in the Department of Human Genetics, Emory University School of Medicine, and a faculty member at the Center for Neurodegenerative Diseases at Emory University. He also serves as the director of Emory Proteomics Service Center. Dr. Peng received his B.S. degree in Biochemistry from Wuhan University in China in 1991, his Ph.D. degree in Biochemistry from University of Iowa in 1999 under the guidance of Dr. David Price. Before joining Emory University in 2002, He performed postdoctoral training in Neuroscience with Dr. Li-Huei Tsai, and in Mass Spectrometry/Proteomics with Dr. Steven Gygi at Harvard Medical School. His research interest is to develop mass spectrometry-based proteomics approaches and to apply the tools to the study of posttranslational modifications and diseases. His current focus is the large-scale analysis of protein ubiquitination and the role of ubiquitination in pathogenesis. He has authored a total of 55 peer-reviewed papers and 11 reviews and book chapters.

### **Richard D. Smith, Ph.D., Battelle Pacific Northwest Laboratories**

Dr. Smith is an adjunct faculty member of the Department of Chemistry, Washington State University, and the Department of Chemistry, University of Utah, and an affiliate faculty member of the Department of Chemistry, University of Idaho. He has presented more than 350 invited or plenary lectures at national and international scientific meetings, and is the author or co-author of more than 600 publications. Dr. Smith holds twenty-seven patents and has been the recipient of seven R&D 100 Awards. He is Director of the NIH Research Resource for Integrative Proteomics located at PNNL.



### **Stephen P. Walton, Ph.D., Michigan State University**

Dr. Walton received his B.ChE. from Georgia Tech, and a M.S. (Chemical Engineering Practice) and Sc.D. in the Department of Chemical Engineering at MIT. While at MIT, he was awarded a Shell Foundation Fellowship and was an NIH Biotechnology Predoctoral Trainee. Upon completion of his Sc.D., he joined the Stanford Genome Technology Center, receiving an NIH Kirschstein postdoctoral fellowship for his research. His research is focused on the application of genomics tools to the measurement of DNA, RNA, and protein expression profiles, as well as the engineering of active biomolecules through kinetic and thermodynamic design.

### **Nathan J. Edwards, Ph.D., Georgetown University**

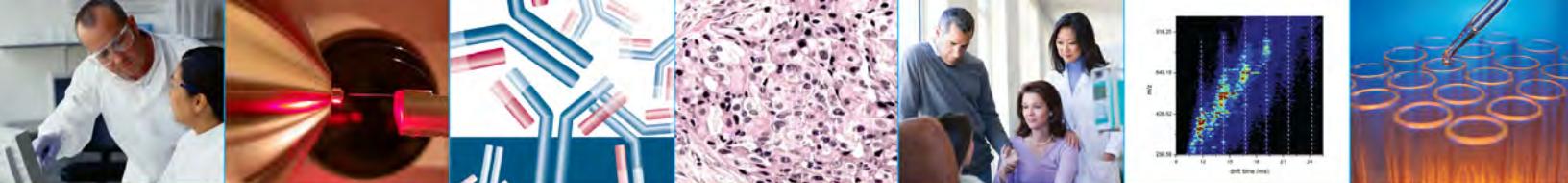
Dr. Edwards received a Ph.D. in Operations Research from Cornell University in 2001. Joining the Informatics Research group at Celera Genomics, Dr. Edwards worked on SCOPE, for identifying peptides from tandem mass spectra by searching protein sequence databases, and other critical elements of the analysis infrastructure for Celera's high-throughput proteomics facility. Moving to Applied Biosystems, still as part of the Informatics Research group in 2002, he led research on algorithmic and statistical issues arising in the analysis of proteomics biomarker workflows and developed the Biomarker Toolbox prototype.

Since joining the Center for Bioinformatics and Computational Biology at the University of Maryland, College Park, in 2004, Dr. Edwards' research has focused on the discovery of novel peptides that characterize alternative splicing, coding SNPs, and mutant protein isoforms, using genomic and EST sequences; and on the rapid identification of microorganisms by MALDI-TOF mass spectrometry and bioinformatics, in collaboration with researchers at University of Maryland, College Park; and the Johns Hopkins School of Public Health. In 2008, Dr. Edwards became an assistant professor in the Department of Biochemistry and Molecular & Cellular Biology at Georgetown University Medical Center.

### **Dariya Malyarenko, Ph.D., College of William and Mary**

Dr. Malyarenko is a technical lead for the collaborative initiative between the College of William and Mary, INCOGEN, Inc., and Eastern Virginia Medical School targeted at developing new algorithms and software tools for signal processing and statistical analysis of mass spectrometry data with intended use for cancer biomarker discovery. Her area of expertise is in experimental and computational physics applied to analysis of nuclear magnetic resonance and mass spectrometry data for synthetic polymers and biomolecules. Her current research interests are in time series analysis, noise filtering, resolution enhancement, and classification of spectroscopic data for cancer profiling proteomics.

### **Denkanikota Mani, Ph.D., Massachusetts Institute of Technology**



Dr. Mani is a Senior Computational Biologist in the Cancer Program and Proteomics Group at the Broad Institute of MIT and Harvard. He is an experienced computational scientist with extensive training and expertise in computer science; pattern recognition and machine learning; data mining; and parallel computing. He has been directly responsible for the design and implementation of pattern-based proteomic data analysis methods for biomarker discovery from a range of mass spectrometric data. The proteomics pipeline he developed has been used to analyze a variety of human and animal tissues representing a spectrum of cancers and to discover biomarker candidates for enabling early detection, diagnosis and targeted therapeutics. Dr. Mani has also applied computational pattern recognition methodology to address biological problems involving diverse data ranging from gene expression profiles to small molecule screens.

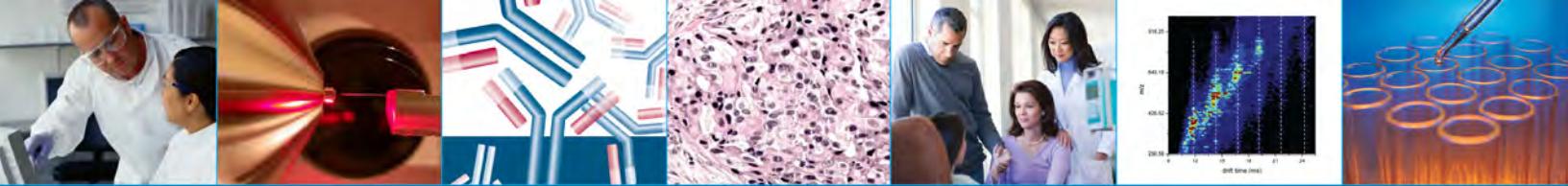
As part of the current program, Dr. Mani and his collaborators at The Broad Institute are working on developing robust, high-throughput methods for proteomic biomarker discovery using high information content mass spectrometry data from state-of-the-art instrumentation. This novel platform combines peptide identity with high-resolution mass spectrometric pattern to identify potential biomarkers comprising both identified peptides/proteins, and statistically significant mass spectrometric peaks and patterns easily amenable to subsequent identification.

Dr. Mani has a Bachelor's degree in Electronics & Telecommunication Engineering from Bangalore University, India, a Master's degree in Computer Science from the Indian Institute of Technology at Kanpur, and a Ph.D. in Computer and Information Science from the University of Pennsylvania at Philadelphia. Prior to joining The Broad Institute, he was a Principal Member of Technical Staff at Verizon Laboratories, where he applied pattern recognition to diverse business and customer data in order to glean insights enabling informed and data-driven corporate decision making. He was also a Principal Research Engineer at Thinking Machines Corporation (now part of Oracle), where he designed and implemented massively parallel data mining algorithms, and applied them to mining large data warehouses.

### Alexey I. Nesvizhskii, Ph.D., University of Michigan

Dr. Nesvizhskii received an M.S. (with honors) from St. Petersburg State Technical University, Department of Physics and Technology, St. Petersburg, Russia in 1995 and a Ph.D. in Physics from the University of Washington, Seattle in 2001. He completed postdoctoral training in Ruedi Aebersold's Lab at the Institute for Systems Biology in Seattle, Washington from 2001-2003 and joined the staff as a Research Scientist upon completion of training.

Dr. Nesvizhskii was the recipient of a medal for "Best Student Scientific Work" awarded by the Russian Federation State Committee of Higher Education and was named Russian Presidential Fellow for the period 1994-1995 and Soros Fellow for the period 1995-1996. In November 2005, Dr. Nesvizhskii joined the faculty of the Department of Pathology as an Assistant Professor.



Dr. Nesvizhskii's research interest is in the field of quantitative proteomics, with a focus on the development of computational methods for processing and extracting biological information from complex proteomic datasets. Similar to other global high throughput technologies such as microarray gene expression analysis, proteomics is extremely dependent on the ability to quickly and reliably analyze large amounts of experimental data. One of the aims of Dr. Nesvizhskii's research is to close the critical gap between the development of high throughput quantitative proteomics methods and the ability to deal with the resulting data deluge and to convert it into new biological knowledge or to develop new disease biomarkers. The efforts in his lab range from the development of computational tools and statistical methods for mass spectrometry-based peptide and protein identification and quantification, to the establishment of guidelines and standards for proteomic data analysis and publication, to the creation of public databases and proteomic data repositories and integration of proteomic with genomic and other types of biological data.

*Awards/Honors:*

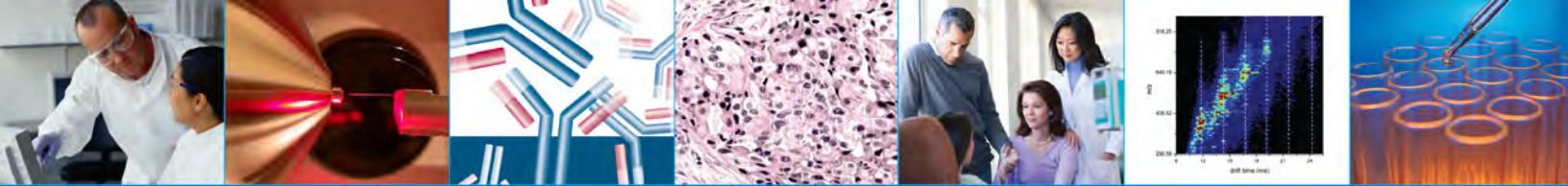
"Rising Young Investigator" by Genome Technology Magazine, Dec. 2006/Jan 2007

**Timothy W. Randolph, Ph.D., Fred Hutchinson Cancer Research Center**

Dr. Randolph is a Senior Fellow at the University of Washington, Department of Biostatistics and a Senior Staff Scientist at the Fred Hutchinson Cancer Research Center's program on Biostatistics and Biomathematics. He is completing a career transition from pure mathematics to one focusing on the analysis of high-dimensional data from proteomic platforms for which he is developing methods for signal processing, statistical analysis, and classification.

Prior to his current positions, Dr. Randolph was an Associate Professor of Mathematics and Statistics at the University of Missouri-Rolla where his research focused on abstract dynamical systems, including the mathematics of control theory. His new research draws on this background using tools from harmonic/wavelet analysis, functional analysis, and operator theory.

Dr. Randolph received his Ph.D. in mathematics from the University of Oregon. While at the University of Missouri, his research was facilitated by several awards from the Missouri Research Board. In 2002, he was awarded a grant from the NIH Institute of General Medical Sciences program on Mentored Quantitative Research Career Development, allowing him to turn his attention to problems related to health and disease. He has authored papers in journals ranging from pure mathematics to computer methods and engineering, to molecular proteomics.



### Katheryn A. Resing, Ph.D., University of Colorado at Boulder

Dr. Resing holds a Ph.D. from the University of Washington, Seattle, and did her postdoctoral fellowship at the University of Washington, Seattle. Her research centers around developing global protein analysis of mammalian cells. The specific technique she uses involves digestion of an extract from cells into peptides, followed by multi-dimensional chromatography of the peptides, where the final stage is reverse phase coupled to a mass spectrometer. The information in this data is then used to search a protein or peptide database, in order to identify the peptides. The use of peptide databases provides more sensitive search results in situation where normal search methods create a huge database, e.g., for splice junctions, alternative start/stop sites, unannotated genes, and modified peptides.

A large part of her research involves developing new methods for data management and mining, including new methods for validating the peptide assignments, and for quantifying and comparing different samples or samples analyzed using different MS instruments and methods. She is also developing new informatics tools utilizing this information in order to tackle several biological and clinical problems, that is, how does the proteome reflect modification of signaling processes in the cell, including expression, splicing, and phosphorylation.

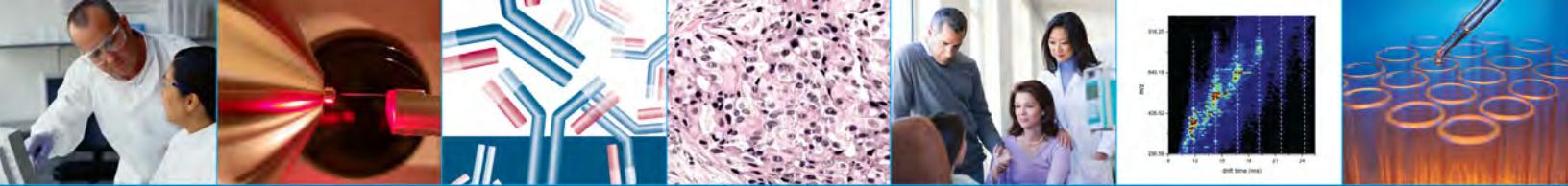
### David L. Tabb, Ph.D., Vanderbilt University

Dr. Tabb was named one of two White House Presidential Scholars for Missouri in 1992. He attended college at the University of Arkansas as a Sturgis Fellow. He majored in Biology and minored in Computer Science, graduating *summa cum laude* in 1996.

He studied proteomics as a graduate student under John Yates at The Scripps Research Institute. Dr. Tabb has created several software tools for data mining proteomic results to extract biological information more consistently and rapidly- e.g., DTASelect, GutenTag software package, DBDigger, and MS2Grouper. Dr. Tabb joined the faculty of Vanderbilt University Medical Center in 2005 to lead a group in the Mass Spectrometry Research Center.

### Dennis J. Templeton, Ph.D., University of Virginia

Dr. Templeton received his M.D. and Ph.D. degrees at the University of California Southern California, and trained in Pathology at the New England Deaconess Hospital. He has long been active in cancer research, earning his Ph.D. at the Salk Institute studying tumor virology, and a postdoc at the Whitehead Institute in molecular oncology and signal transduction. After ten years at Case Western Reserve University, he was appointed Chair of Pathology at the University of Virginia. As part of the mission of pathology to discover improved diagnostics, he established the Pathology MS Facility and the biomarker discovery program, and has been closely involved in the day-to-day work, particularly that involving peptide chemistry.



## 4.3 Responsibilities of CPTAC Members

### Program Coordinating Committee

#### Chair

- The primary responsibilities of the Chair include general oversight of Assembly activities.
- The Chair ensures that all reports (CPTAC Teams and the PCC) are delivered to the NCI in a timely manner.
- Meetings
  - Planning meeting agendas with CPTAC Team Leaders and the NCI.
  - Distributing the meeting agenda to the PCC.
  - Ensuring that all CPTAC meetings are meaningful and productive.
  - Determining the scientific vision for meeting, in coordination with the PCC Co-Chair and the NCI.
  - Ensuring the quality of the Technical Program.
  - Developing teleconference schedule, in coordination with the PCC Co-Chair, and the NCI.
  - Ensuring that the appropriate text is provided for Advance and Final Programs. This includes all session scheduling, presentation, and verified spelling of presenters' names and affiliations. To be coordinated with the PCC Co-Chair and the NCI.

#### Co-Chair

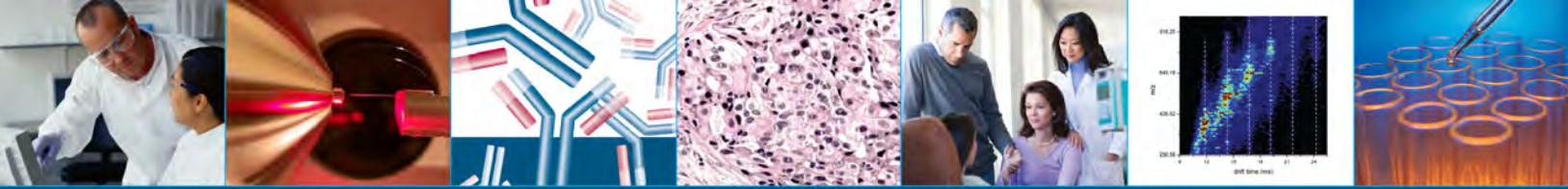
- The primary responsibility of the Co-Chair is to assist the Chair, and prepare to serve as the next Chair.
- The Co-Chair is responsible for oversight of ad-hoc representatives to the PCC and Working Groups.
- Presides at all PCC meetings and to co-chair facilitating meetings.
- Enforcing attendance standards (if a member has 2 unexcused absences the co-chair will appoint another member to serve on the committee).

#### Voting Members of the PCC

- The primary responsibility of the other members of the PCC is to make decisions that advance the mission of the CPTAC program.
- Attend meetings and teleconferences.
- Implement decisions of the PCC across the team which they lead.

#### Ad Hoc Members of the PCC

- The primary responsibility of ad hoc members of the PCC is to advise the voting members in areas of their expertise.



## Working Group

### Chair

- The primary responsibility of the Working Group Chair is to carry out the mission of the Working Group.
- Calls meetings, plans meeting agenda, and facilitates meetings.
- Reports recommendations, action items and milestones to the PCC.
- Alerts the PCC to challenges arising within the Working Group.
- Participates in a monthly Working Group Chairs call, intended to coordinate efforts between working groups.

### NCI Point-of-Contact

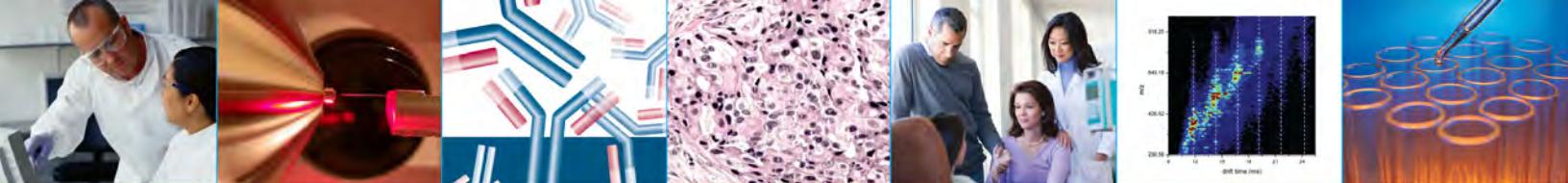
- The primary responsibility of the NCI Point-of-Contact is to provide the Working Group with oversight, ensuring that the Working Group is implementing the directives of the PCC.
- Serve the Working Group by answering questions regarding programmatic details pertinent to that Working Group.
- Keep the NCI informed of the goings-on of each working group.

## NCI Organizational Team

- The NCI Organizational Team shall hold other CPTAC members to agreed-upon programmatic goals.
- Interact regularly with each component of the program so as to coordinate efforts across the program.
- Explore outreach opportunities to raise awareness of the CPTAC program and foster strategic partnerships with the broader scientific community.
- Deliver program-wide progress reports to senior NCI staff and advisory committees.

## NCI Point-of-Contact to a CPTAC Center

- The primary responsibility of the NCI Point-of-Contact is to provide a CPTAC Center with oversight, ensuring that the Center is implementing the decisions of the PCC.
- Serve as conduit between a CPTAC Center and the NCI for ensuring that programmatic milestones are met.
- Serve the Center by answering questions regarding programmatic details pertinent to that Center.
- This individual is an NCI employee.

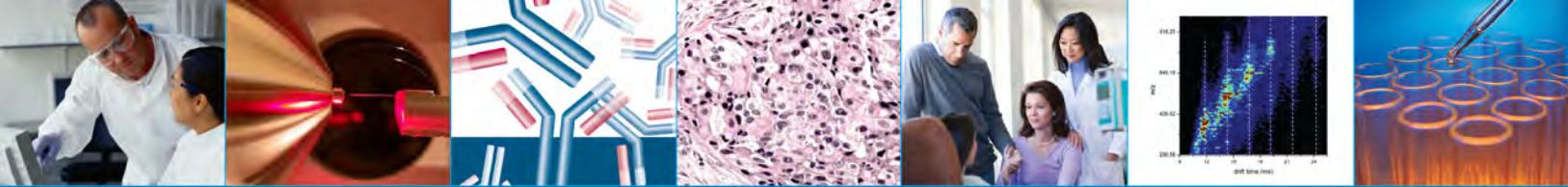


## NCI Partners

- NCI partners provide expertise in a number of specialty areas: standards, reagents, statistical design, event planning, etc.
- NCI partners will advise and aid the NCI according to their expertise.
- On occasion, the NCI may recommend a partner as an ad hoc member of the PCC or Working Group. Such a recommendation will then be subject to the rest of the PCC.
- Unless authorized by the NCI, NCI partners do not officially communicate with the CPTAC centers.

## External Oversight Committee

- Members of the External Oversight Committee provide visionary advice on the overall direction of the CPTAC program.
- Composed of high-level individuals throughout the Life Sciences, the committee will counsel the PCC on how best to interface with other fields, take advantage of new opportunities, or proceed with regulatory approval.
- The PCC presents an annual report to the External Oversight Committee.



## 4.4 Program Coordinating Committee Biographies

### PCC Members

#### **Steve Carr, Ph.D., Broad Institute of MIT and Harvard**

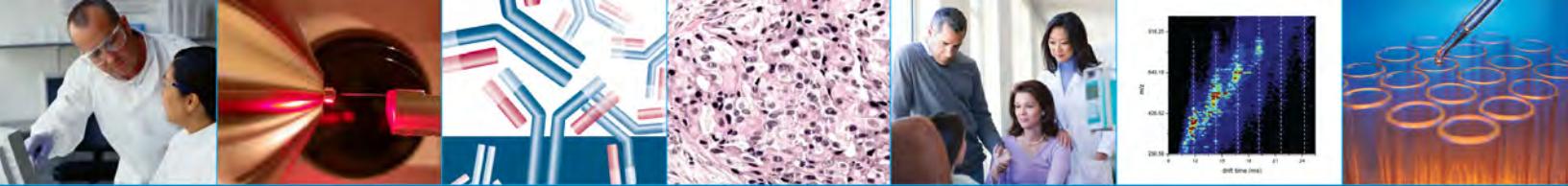
Dr. Carr leads the Proteomics and Biomarker Discovery efforts at the Broad Institute of MIT and Harvard. His current research involves development and application of novel, quantitative approaches for biomarker discovery and validation as well as phosphor-proteomics to understand drug and disease mechanisms and provide biomarkers for human cancers, heart disease and metabolic disorders. The Carr lab collaborates with biologists and chemists to systematically identify proteins and their modifications whose abundance or form is modulated by disease or drug action, as well as to define physical and functional associations of protein constituents of regulatory and signaling pathways involved in health and disease. These studies involve analysis of complex biological specimens, such as tumor tissues or patient blood using protein chemistry and advanced separation methods together with state-of-the-art mass spectrometry.

For the last 25 years, Dr. Carr's research has focused on applying and developing proteomics methods in order to understand the mechanism of action of drug candidates and build an understanding of protein targets and their roles in disease. He is noted for developing methods for selective enrichment, detection and quantitation of posttranslational modifications such as phosphorylation and glycosylation in the proteome. While at GlaxoSmithKline (1984-2001) and Millennium Pharmaceuticals (2001-2004), he made significant contributions to the discovery and development of four marketed drugs (small and large molecule) and to numerous drugs in clinical trial. His groups at GSK and Millennium also produced one of the first examples of proteomics-derived biomarkers in use in the clinic and the first example of de-orphaning of a G-coupled protein receptor by mass spectrometry. He has over 130 peer-reviewed publications on development and use of proteomics and biological mass spectrometry. Dr. Carr is an Associate Editor of Molecular and Cellular Proteomics, and he has served on the editorial boards of Analytical Chemistry and Protein Science, among others.

Dr. Carr received his B.S. in 1976 from Union College and Ph.D. from MIT in 1980. After four years of postdoctoral training at Harvard Medical School and MIT, he joined SmithKline Pharmaceuticals (now GlaxoSmithKline), becoming director of Computational and Structural Sciences in 1997. Most recently he led protein science and proteomics groups at Millennium Pharmaceuticals in Cambridge, MA, prior to joining the Broad Institute in 2004.

#### **Susan Fisher, Ph.D., University of California, San Francisco**

Susan Fisher, Ph.D., is currently a Professor of Cell and Tissue Biology and the Faculty Director for the Biomolecular Resource Center Mass Spectrometry Center. Her research is focused on many aspects of reproductive biology as well as the biomedical applications of mass spectrometry. She has received numerous honors throughout her career, including an achievement award from the Society for Women's Health. Fisher



received her Bachelor degree from Hope College in Holland, Michigan, and her Ph.D. degree from the University of Kentucky, where she also completed a postdoctoral fellowship in mass spectrometry.

### **Dan Liebler, Ph.D., Vanderbilt University**

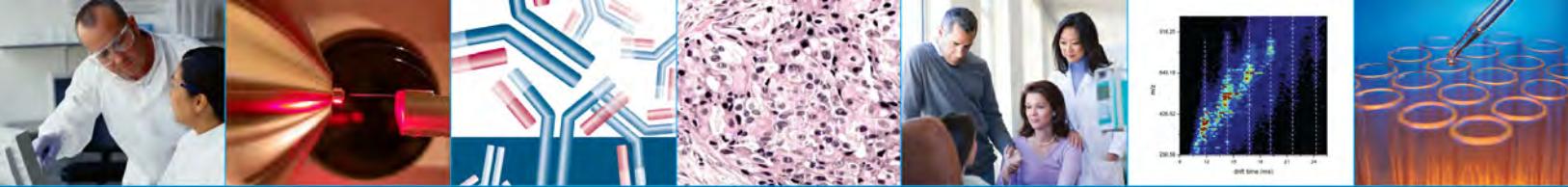
Dr. Daniel C. Liebler received a Ph.D. in Pharmacology from Vanderbilt while training in the laboratory of F. Peter Guengerich. He did postdoctoral training in the laboratory of Donald J. Reed at Oregon State University. From 1987-2003, Dr. Liebler served on the faculty of the Department of Pharmacology and Toxicology in the College of Pharmacy at the University of Arizona. From 1987 to 1998, his research program focused on the mechanisms of action of antioxidants in prevention of oxidative damage and cancer. His group defined antioxidant pathways and mechanisms of antioxidant action of vitamin E, carotenes and flavonoid antioxidants and established the use of mass spectrometry (MS)-based assays for products of antioxidant reactions as markers of antioxidant function. Since 1998, Dr. Liebler's program has focused on the application of MS-based proteomics approaches to address the problem of how reactive chemical intermediates cause damage to proteins.

The Liebler group developed two data analysis tools, SALSA and P-Mod, which enable discovery of modified protein forms through analysis of MS data, even when the chemical nature and sequence specificity of the modifications are not known beforehand. This approach has been integrated with affinity capture methods to evaluate electrophile-mediated protein damage on a proteomic scale and map over 1500 sites of chemical modification on over 800 human proteins. Dr. Liebler's laboratory has also extended these proteomic approaches to analyze the role of site-specific modifications in the function of signaling and sensor proteins that regulate responses to chemical toxicity and oxidative stress. The mapping of modifications has been augmented by application of stable isotope tagging methods to analyze the kinetics of protein modification reactions.

Dr. Liebler relocated to Vanderbilt in June 2003, where he has served as Director of Proteomics in the Mass Spectrometry Research Center. The Proteomics Laboratory has implemented methods and approaches developed in Dr. Liebler's laboratory and serves over 130 research groups at Vanderbilt. Most recently, Dr. Liebler has accepted Directorship of the Jim Ayers Institute for Precancer Detection and Diagnosis, which is dedicated to the discovery of proteomic markers for early cancer detection and for guiding therapy of established disease. Dr. Liebler's long-term research goals are to apply proteomics and related emerging technologies to identify markers of disease, therapeutic effect and toxicity and to characterize the roles of protein damage in chemical toxicity and disease.

### **Paul Tempst, Ph.D., Memorial Sloan-Kettering Cancer Center**

Dr. Tempst leads a proteomics research team specializing in R&D of protein / peptide microanalysis, with applications in chromatin dynamics research and cancer biomarker discovery. He is a Member of the Sloan-Kettering Institute and Professor at the Gerstner Sloan-Kettering Graduate School of Biomedical Sciences, as well as Professor of



Molecular Biology at the Weill Graduate School of Medical Sciences, Cornell University. He provides executive oversight of the microchemistry, proteomics, genomics and engineering resource laboratories at his host institution. He has over 25 years of experience in protein chemistry and biochemistry, and mass spectrometry, and he has collaborated with investigators worldwide to identify and characterize novel and unknown interacting proteins and protein complexes.

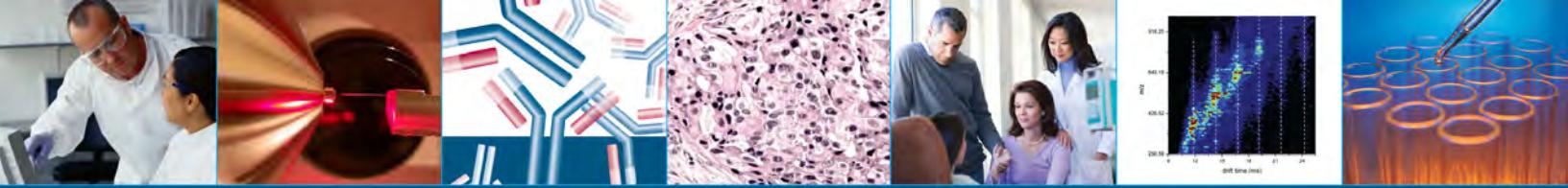
His research is documented in over 230 peer-reviewed papers on proteomics methods and applications, mostly to study the mechanisms of transcriptional regulation. He is an inventor on several patents, including microfluidics-based automated chemistries, a continuous nano-electrospray ion source for mass spectrometry, Gallium-affinity chromatography of phosphopeptides, a protein-nitrosylation assay, and antibody micro-arrays with multi-layered affinity detection. Recently, his lab developed a novel platform for serum peptidome analysis and advanced the model that cancer type-specific peptide patterns are surrogate markers for changing exoprotease panels.

Dr. Tempst obtained a B.S. (1976) and Ph.D. (1981) at Ghent University in his native Belgium. He received a NATO Research Fellowship to conduct postdoctoral studies (1982-85) in the laboratory of Dr. Leroy Hood at the California Institute of Technology, where he was involved in the development of the earliest microchemistry systems. He later received a Leukemia Society of America Special Fellow Award and the Irma Hirsch Trust Career Scientist Award for the application of these novel technologies in molecular biology, genetics and medicine. He was a faculty member at the Harvard Medical School (1986-90) before taking up his current positions in New York in 1991.

### **Fred Regnier, Ph.D., Purdue University**

After completing a Ph.D. at Oklahoma State University (1965) and Post Doctoral work at the University of Chicago (1966) and Harvard (1968), Professor Regnier became an Assistant Professor of Biochemistry (1968) at Purdue University. During twenty years in Biochemistry at Purdue he was an Associate Professor (1971-76), Professor (1976-90) and Associate Director of the Agricultural Experiment Station (1976-78). He became a Professor of Chemistry in the Chemistry Department at Purdue in 1990 and was promoted to Distinguished Professor in 2004. He is the author of over 300 publications, 40 patents, and several books on various aspects of chemistry, biochemistry, and particularly separation science and immunological assays as they relate to proteins.

Professor Regnier has for several decades been involved in the transfer of technology to society through patents from his laboratory, company creation, and continuing education. Along with Professor Barry Karger (Northeastern University), he co-founded Bioseparations in 1984 to provide advanced courses in analytical chemistry to scientists in the pharmaceutical industry. During the course of the next four years Bioseparations presented courses to more than a thousand scientists in U.S and European pharmaceutical companies. He then co-founded PerSeptive Biosystems (PBIO) with Noubar Afeyan in 1988 based on licensed Purdue technology, and actively participated in the growth of that company to 500 people in the U. S., Europe, and Asia. The sale of large-scale pharmaceutical manufacturing systems and scientific instruments by PBIO to



more than 300 pharmaceutical and biotechnology companies grew to \$100 million in 1997, when the company was sold to Applied Biosystems (ABI). In 2000, he co-founded Beyond Genomics (BG) with Jan van der Greef (University of Leiden), David Clemmer (Indiana University), and Scott McLuckey (Purdue University). BG provides advanced drug discovery tools to the pharmaceutical and biotechnology industry. The company has grown to 60 people with facilities in Europe and North America and has ongoing drug discovery relationships with major pharmaceutical companies. Most recently, he co-founded Quadraspec with David Nolte, Chard Barden, and Eric Davis. The focus of Quadraspec is on very high throughput immunological assay technology for health assessment.

### **Henry Rodriguez, Ph.D., M.B.A., National Cancer Institute**

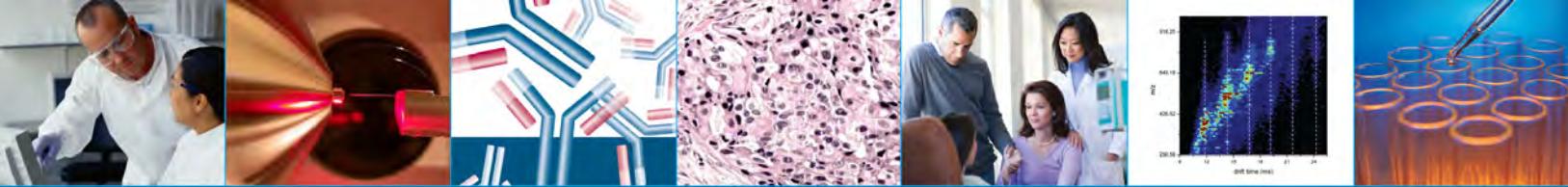
Dr. Rodriguez serves as the Director of the Clinical Proteomic Technologies for Cancer programs within the NCI's Office of Technology and Industrial Relations, in the Office of the Director. In this role, he oversees and is responsible for the vision, direction, and implementation of highly complex scientific proteomic technology programs. These initiatives consist of the Clinical Proteomic Technologies for Cancer initiative (CPTC) and the Mouse Proteomic Technologies Initiative (MPTI).

Dr. Rodriguez is an internationally recognized expert in advanced molecular-based cancer technologies, specifically proteomics, genomics, cellomics, and bioinformatics. Immediately before coming to the NCI in 2006, Dr. Rodriguez was at the NIST, where he held several roles (1998-2006). At NIST, Dr. Rodriguez developed and was the Leader of the Cell and Tissue Measurements Group, where he successfully established four advanced technology research programs in the areas of quantitative cell biology/cellomics, proteomics, gene expression, and bioinformatics. Initiatives in proteomic metrology included protein capture chemistry, mass spectrometry, protein sequencing, peptide synthesis, and database development.

Dr. Rodriguez also developed a tissue engineering program, of which he was the program manager and chairman of the NIST Regenerative Medicine Strategic Working Panel. In this capacity, he established molecular-based analytical procedures to measure genetic damage in tissue-engineered medical products that may result from manufacturing, storage or shipping, and coordinated these efforts with the US Food and Drug Administration (FDA) and ASTM International.

Dr. Rodriguez helped pioneer new measurement methodologies that utilized LC/MS. He is internationally known for his outstanding efforts in developing, applying, and quantifying oxidatively-modified base detection methods for applications in health care and cancer diagnostics. Dr. Rodriguez has been awarded several Department of Commerce ATP intramural awards and served as an ATP Proposal Scientific Expert.

Dr. Rodriguez served as a Program Analyst (Science/Strategic Policy Expert) in the Office of the Director at NIST, where he helped in the planning and evaluation of strategic scientific policies and partnerships in biotechnology, including interactions with the Department of Commerce and members of Congress. In 2005, Dr. Rodriguez



developed the *Measurement Challenges in Proteomics* workshop, as part of NIST's Roadmap Initiative. He also participated on Department of Commerce efforts in education and/or training in genetics and genetic technologies of professionals for the Secretary's Advisory Committee on Genetics, Health, and Society.

Dr. Rodriguez is an advocate that molecular medicine is transforming the processes of drug discovery, development, and delivery, suggesting that the war on cancer is embarking upon a new era. By providing a suite of rapid, sensitive, specific, high throughput analytic tools and supporting standardized reagents and informatics capabilities, the integrated application of advanced technologies offers the possibility of transforming this challenge into an opportunity.

Dr. Rodriguez began his career as a fellow at the Department of Immunology of The Scripps Research Institute, and then at the Department of Medical Oncology of the City of Hope National Medical Center. Dr. Rodriguez is the recipient of domestic and international awards, including the Sigma Xi Young Scientific Investigator award, the Alumni Honor Roll of the MARC/MBRS programs of the National Institute of General Medical Sciences at the National Institutes of Health, a Proclamation letter signed by the Governor of Hawaii, and the Science Spectrum Trailblazer, Top Minority in Research Science Award. In addition, he has been elected to the presidency of scientific organizations such as the Oxygen Club of Greater Washington, D.C. He has authored over 59 papers in peer-reviewed journals and books and co-edited a science book titled Oxidative Stress and Aging: Advances In Basic Science, Diagnostics and Intervention, that made the publisher's "Bestsellers List."

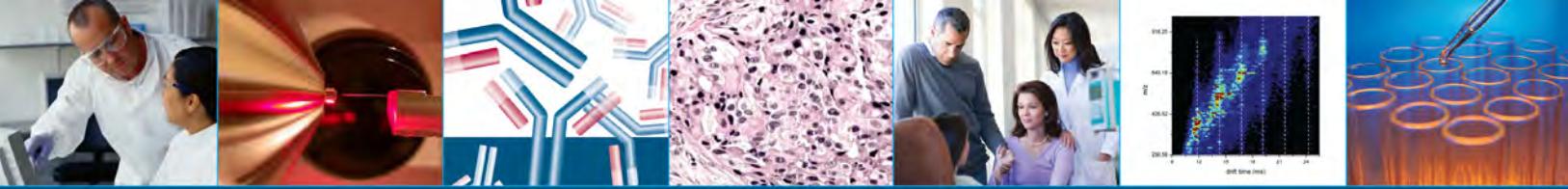
Dr. Rodriguez received a B.S. and M.S. in Biology from Florida International University, a Ph.D. in Cell and Molecular Biology from Boston University, and an M.B.A. from Johns Hopkins University School of Business Management.

### Ad-hoc PCC members

#### **Leigh Anderson, Ph.D., Plasma Proteome Institute**

Leigh Anderson has served as the Chief Executive Officer of the Plasma Proteome Institute, a scientific research institute in Washington, D.C., of which he is also a founder (since 2002). Dr. Anderson also consults through Anderson Forschung Group, of which he has been a Principal since 2002 and a member of its board of directors since 2004. From 2001 to 2002, Dr. Anderson served as the Chief Scientific Officer and a member of the board of directors of Large Scale Biology Corporation, a biotechnology company that previously traded on NASDAQ under the symbol "LSBC." Dr. Anderson also served as a member of the board of directors and a member of the audit committee of Dade Behring Holdings, Inc. (DADE), a NASDAQ-listed company, from 2002 until its acquisition by Siemens AG in 2007. Dr. Anderson earned a B.A. in Physics from Yale University and a Ph.D. in Molecular Biology from Cambridge University.

#### **Bradford W. Gibson, Ph.D., Buck Institute of Age Research**



Brad Gibson is currently Professor and Director of Chemistry at the Buck Institute of Age Research, an independent non-profit research institute founded in 1999 to study the basic biology of aging and age-related disease. Since 1985, Dr. Gibson has also been a faculty member at the University of California, San Francisco (UCSF) in the Department of Pharmaceutical Chemistry. Prior to his work at the Buck Institute and UCSF, Brad received his Ph.D. in Klaus Biemann's lab at Massachusetts Institute of Technology where he developed mass spectrometry approaches to study protein primary structures, and then as a postdoctoral fellow in Dudley Williams group at Cambridge University to study bioactive peptides. While at the Buck Institute, Dr. Gibson has focused his research interests on advancing proteomic technologies to aging research, with a focus on mitochondria, oxidative damage, and protein phosphorylation. More recently, he has expanded these efforts to include studies in protein aggregation, turnover and disease-specific biomarker discovery.

#### **Joe Gray, Ph.D., Lawrence Berkeley National Laboratory**

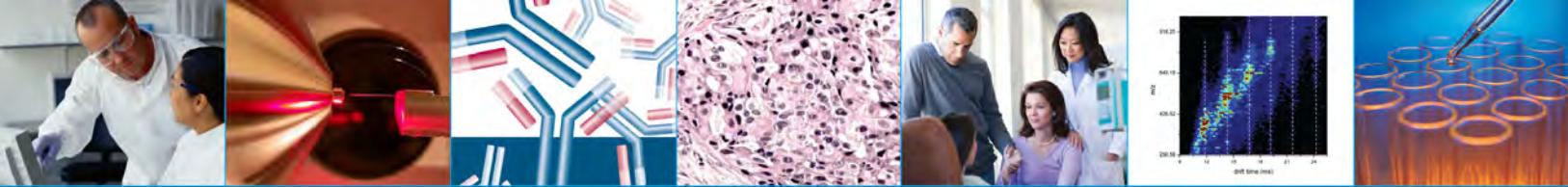
Dr. Joe Gray majored in engineering physics at the Colorado School of Mines and obtained his Ph.D. in Nuclear Physics from Kansas State University in 1972. He then joined the Biomedical Sciences Division of the Lawrence Livermore National Laboratory, moving to UCSF as Professor of Laboratory Medicine and Radiation Oncology in 1991. He established and headed the Division of Molecular Cytometry in the Department of Laboratory Medicine until 1997. He was Interim Director of the UCSF Cancer Center from 1995 to 1997 and is now Program Leader for Cancer Genetics and Breast Oncology there. He has been Principal Investigator of the Bay Area Breast Cancer SPORE since 1996. Dr. Gray accepted a position as Division Director of Life Sciences and Associate Director of Biosciences at the Lawrence Berkeley National Laboratory in April 2003, and will continue as a member of the UCSF Cancer Center and as Principal Investigator of the Breast Cancer SPORE.

#### **Lee Hartwell, Ph.D., Fred Hutchinson Cancer Research Center**

Dr. Leland Hartwell is President and Director of Seattle's Fred Hutchinson Cancer Research Center and Professor of Genome Sciences at the University of Washington. Dr. Hartwell is a member of the National Academy of Sciences and has received the Albert Lasker Basic Medical Research Award, the Gairdner Foundation International Award, the Alfred P. Sloan Award in Cancer Research, and the 2001 Nobel Prize in Physiology or Medicine.

#### **Gordon Mills, M.D., Ph.D., M.D. Anderson Cancer Center**

Dr. Mills is Chair of the Department of Molecular Therapeutics, the Ann Rife Cox Chair in Gynecology, and Professor of Medicine and Immunology at the University of Texas M.D. Anderson Cancer Center. He currently serves on a number of research and advisory boards at the University of Texas and the University of Toronto. Dr. Mills is also on the external advisory board of the Seattle ovarian cancer Specialized Program of Research Excellence, a program of the National Cancer Institute that promotes the transfer of cancer research into the clinical setting. His interests in ovarian cancer include the prognosis and discovery of therapeutic targets and treatments of the disease. His contributions include developing a potential marker for early-stage ovarian cancer, and



studying this as a possible target for therapy. He has also studied genetic mutations in the phosphatidylinositol-3 kinase/PTEN/AKT pathway as another possibility for therapy.

#### **Amanda G. Paulovich, M.D., Ph.D.**

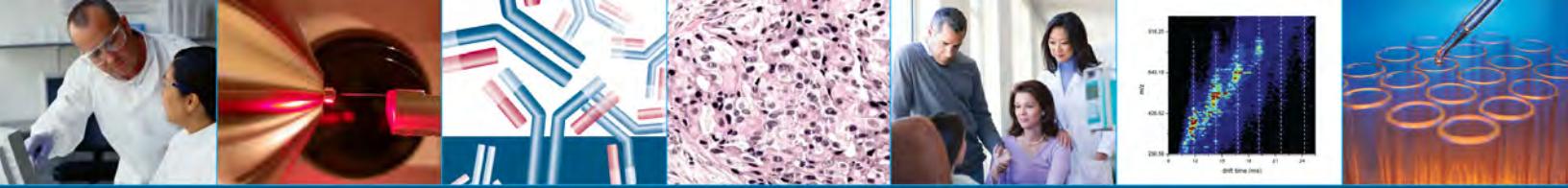
Mandy Paulovich is the Director of the Fred Hutchinson Cancer Research Center's Early Detection Initiative. With training in both clinical medicine and basic science, she currently manages the efforts of an interdisciplinary team to develop molecular imaging tools, design improved diagnostic tests, and identify risk biomarkers. Mandy earned her BS in biological sciences at Carnegie Mellon University, her Ph.D. in genetics at the University of Washington, and her M.D. at the University of Washington. Before joining the Hutch, she completed her internal medicine residency at Massachusetts General Hospital, her oncology fellowship at the Dana-Farber Cancer Institute, and postdoctoral training in computational biology at the Whitehead Institute's Center for Genomics Research, Massachusetts Institute of Technology.

#### **David Ransohoff, M.D., University of North Carolina Lineberger Comprehensive Cancer Center**

Dr. Ransohoff's primary research interest is in improved methods of colon cancer screening. He has published extensively on the use of colonoscopy screening and surveillance, fecal occult blood testing, sigmoidoscopy, and virtual colonoscopy; his most recent work concerns stool DNA testing, serum proteomics, and the development and assessment of other new 'omics' methods to screen for cancer. Working with NCI's EDRN (Early Detection Research Network) and others, he is leading a multi-center study about serum proteomics to diagnose colon cancer. At UNC he has directed the K30 faculty development program since its inception in 1999, helping train junior faculty build careers in clinical and translational research.

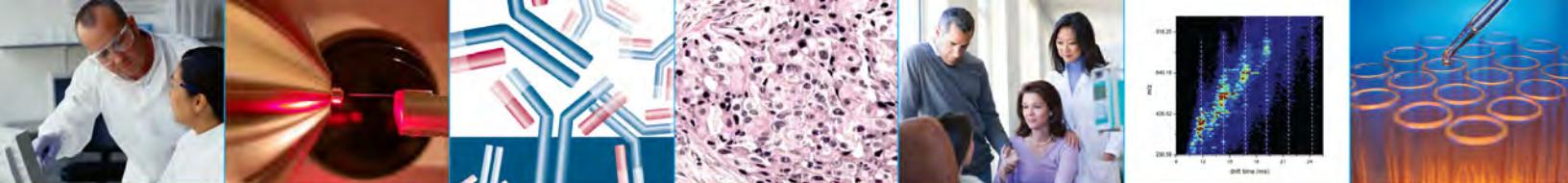
#### **Steven J. Skates, Ph.D.**

Steve Skates is a biostatistician at Massachusetts General Hospital. His research interests focus on early detection of cancer, with a particular emphasis on ovarian cancer. This includes development of algorithms for longitudinal biomarker screening, identification of new biomarkers, and design of screening studies. Steve received his BSc in mathematical sciences at University of Western Australia, and his Ph.D. in Statistics from the University of Chicago.

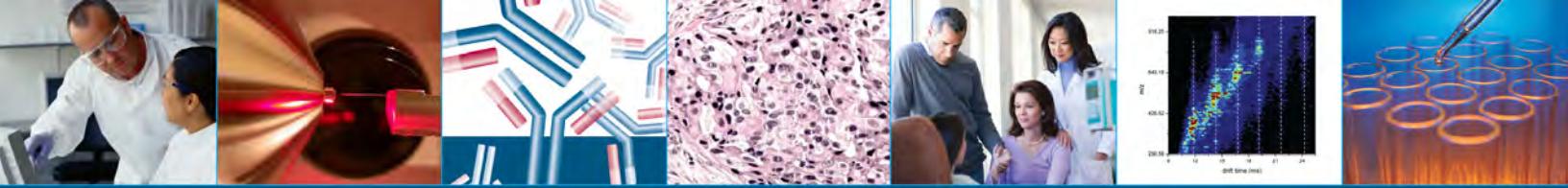


## 4.5 CPTAC Working Group Membership

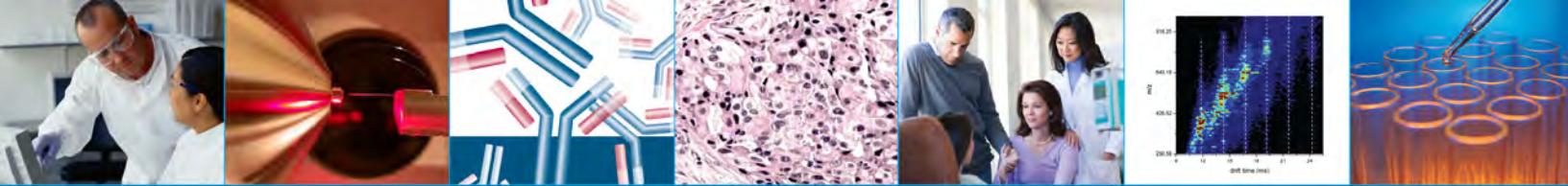
Year 1 Working Group	Members
<b>Experimental Design and Statistics: Unbiased Discovery Studies</b>	Ruth Pfeiffer, <b>Mandy Paulovich*</b> , Dan Liebler, Dean Billheimer, Dave Duewer, Pei Wang, and Cliff Spiegelman, Chris Kinsinger
<b>Experimental Design and Statistics: Verification Studies</b>	Steve Skates, (Pei Wang), Cliff Spiegelman, <b>Steve Hall*</b> , Fred Regnier, Changyu Shen , Dave Duewer, Mu Wang, Leigh Anderson, Adam Clark, Dean Billheimer
<b>Yeast Cell Lysate</b>	<b>Mandy Paulovich*</b> , Dan Liebler, Ron Blackman, Fred Regnier, Henry Rodriguez
<b>Cell Line Selection and Lysates</b>	<b>Joe Gray*</b> , Susan Fisher, Carlos Artiaga, Paul Tempst (or designate), Adam Clark
<b>Plasma Standard Pool</b>	<b>Leigh Anderson*</b> , Terry Pearson, Adam Clark, Dave Bunk, Chris Kinsinger
<b>Protein Standards: Selection and Production</b>	Joe Gray, <b>Steve Carr*</b> , Susan Fisher, Lee Makowski, Fred Regnier, David Bunk, Dan Liebler, Paul Tempst, Adam Clark, Leigh Anderson
<b>Protein Standards: Post-Translational Modification Needs and Production</b>	<b>Brad Gibson*</b> , Susan Fisher, Ewa Witkowska, Steve Hall, Fred Regnier, Dan Liebler, Steve Carr, Henry Rodriguez, Lee Makowski, Fred Regnier
<b>Data Analysis, Storage and Dissemination</b>	Jake Chen, <b>Ron Beavis*</b> , Chris Kinsinger, John Philip, Karl Clauser, Dave Tabb, Liming Yang, Steve Stein
<b>Biospecimen Collection</b>	Helen Moore, Paul Tempst, <b>Steve Skates*</b> , Ron Beavis, Chris Kinsinger, David Ransohoff. Also clinical point of contact for each group: Peggy Porter, Cher Carlisle, Jake Vinson, Laura Esserman



Year 2 Working Group	Members
<b>Experimental Design and Statistics: Unbiased Discovery Studies</b>	<b>Dan Liebler*</b> , Mandy Paulovich, Dean Billheimer, Pei Wang, Cliff Spiegelman, Chris Kinsinger, Mehdi Mesri, Amy Ham, Lisa Zimmerman, Dave Bunk, D.R. Mani, Ron Blackman, Steve Stein, Mu Wang, Karl Clauer
<b>Experimental Design and Statistics: Verification Studies</b>	<b>Steve Hall*</b> , Steve Skates, Dave Bunk, Tom Neubert, Lisa Zimmerman, Pei Wang, Cliff Spiegelman, Fred Regnier, Mu Wang, Leigh Anderson, Dean Billheimer, Mehdi Mesri, Chris Kinsinger, David Ransohoff, Terri Addona, Simon Allen, Ron Blackman, Christoph Borchers, Steve Carr, Nathan Dodder, Brad Gibson, Jason Held, Angela Jackson, Dan Liebler, Amanda Paulovich, Birgit Schilling, Derek Smith, Jeff Whiteaker
<b>Protein Standards: Post-Translational Modification Needs and Production</b>	<b>Brad Gibson*</b> , Susan Fisher, Ewa Witkowska, Steve Hall, Fred Regnier, Dan Liebler, Steve Carr, Henry Rodriguez, Lee Makowski, Leigh Anderson, Bensheng Li, James Ault, Chris Benz, Mehdi Mesri, Chris Kinsinger, Birgit Schilling, Paul Tempst
<b>Protein Standards: Selection and Production</b>	<b>Leigh Anderson*</b> , Joe Gray, Steve Carr, Susan Fisher, Lee Makowski, Chris Benz, Hans Lilja, Birgit Schilling, Fred Regnier, David Bunk, Dan Liebler, Paul Tempst, Mehdi Mesri, Chris Kinsinger
<b>Data Analysis, Storage and Dissemination</b>	<b>Dave Tabb*</b> , Jake Chen, Phil Andrews, Ron Beavis, Chris Kinsinger, Mehdi Mesri, John Philip, Karl Clauer, Liming Yang, Paul Rudnick, Susan Fisher, Rich Niles, Lorenzo Vega Montoto
<b>Biospecimen Collection</b>	<b>Steve Skates*</b> , Helen Moore, Paul Tempst, Chris Kinsinger, Susan Fisher, Marc Consentino, Mehdi Mesri, David Ransohoff, Susan Fisher, Stig Kreps, Linnette Lay, Marc Lim, Nicole Lockhart, Mark Consentino, Mark Robson, Barbara Stein, Christopher Sweeney, Jim Vaught, Josep Villanueva. Also clinical



	point of contact for each group: Peggy Porter, Melinda Sanders, Jake Vinson, Laura Esserman.
<b>Analyte Selection</b>	<b>Mandy Paulovich*</b> , Pei Wang, ChenWei Lin, Susan Fisher, Joe W Gray, Mike Gillette, Ron Blackman, Daniel Liebler, Josep Villanueva, Mark Robson, David Ransohoff, Steve Skates, Fred Regnier, Mehdi Mesri, Chris Kinsinger, David Bunk
<b>Digestion Procedures</b>	David Bunk, Paul Tempst, <b>Amy Ham*</b> , Ashraf Madian, Birgit Schilling, Dan Liebler, Fred Regnier, Ewa Witkowska, Christopher Kinsinger, Leigh Anderson, Lei Zhao, Mehdi Mesri, Mousumi Ghosh, Paul Rudnick, Amanda Paulovich, Steve Skates, Steve Stein, Steven Carr, Steven Hall, Susan Abbatiello, Susan Fisher, Tara Hiltke, Jeffrey Whiteaker, lisa.zimmerman



## 4.6 Site Visit Guidelines and Checklist

The following site visit guidelines are provided to ensure a productive and uniform site visit for all CPTAC centers.

The purpose of the visit is to enable the CPTC program officers to evaluate the organization that has been awarded a cooperative agreement, or contract with stated deliverables. Visits also allow time for Team Leaders (TLs), investigators and Program to discuss and resolve issues associated with an awarded project. Additionally, site visits provide an opportunity to get to know CPTAC members (including junior staff, post-docs, and graduate students) and to become familiar with research facilities. Overall, site visits are designed for the CPTC to help the CPTAC center successfully meet the objectives of the program (stated below).

Each CPTAC center was awarded based on their proposal to meet the following objectives:

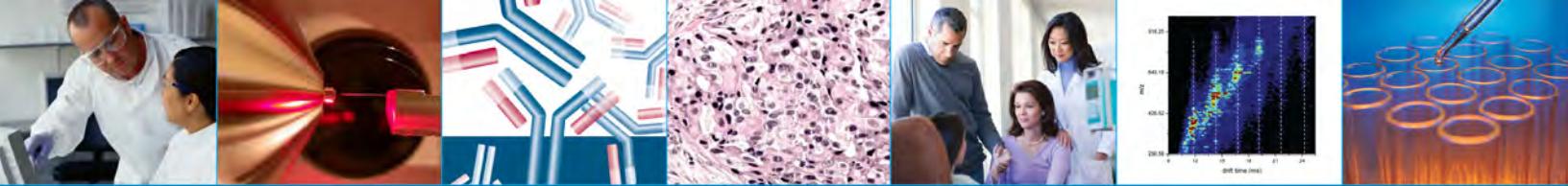
- Evaluate performance of proteomic technology platforms and standardize approaches to developing applications using these platforms;
- Evaluate proteomic platforms for their ability to analyze cancer-relevant proteomic changes in human clinical specimens;
- Establish systematic ways to standardize proteomic protocols and data analysis among multiple laboratories;
- Develop and implement uniform algorithms for sharing bioinformatics and proteomic data and analytical/data mining tools across the scientific community;
- Develop well defined and comprehensively characterized sets of standard/reference materials and reagents to serve as resources for the research community.

### Site Visit Procedure

Program:

- Identifies a site visit team, which normally includes program staff and possibly other staff whose expertise would be beneficial.
- Makes arrangements with CPTAC center TL to identify the number of sites deemed appropriate for stated visit.
- Works with grants management to provide grantees a list of questions to be addressed, in the occurrence of outstanding financial issues.
- Meets with TL, investigators, and staff to discuss progress, challenges, and other issues related to a grant and meets with research administrators to discuss other grant-related issues.

The Center TL (or designate) will coordinate and manage the site visit. A typical site visit should be seven to eight hours in length and focus on the internal projects of the center.



## **Scientific Session**

After introductions, the first item on the agenda is generally a short overview presented by the TL with the key scientific staff present. This presentation should include long-term overview of the project as well as the specific aims for that year. Technical presentations by other scientific staff should follow.

Here is a suggested general format for the scientific session:

- Background/results detail (realization of specific aims, progress toward milestones, updates on the programmatic objectives as stated above)
- Anticipated plans for next year
- Site-specific issues
- Issues related to CPTAC experiments/program

## **Lunch Session**

Lunch should be planned to last less than an hour. It is used most productively as a session to review the preceding presentations, discuss problem areas, and recognize the contributions of the team and core team members. A cold lunch served in the meeting room has been found to be preferable, as it permits maximum flexibility in timing. Visitors are expected to pay for their share of the lunch in advance.

## **Tour of Facilities**

Facility tours are typically brief in duration. Where relevant, they may include walking through important laboratory facilities or clinical units. In general, the best tour times are either at the end of the site visit or immediately around lunch. For example, program officers might choose to consolidate time by touring a facility that is en route to another area where lunch will be held. The tour may be conducted earlier, however, where a visit to a laboratory, for example, is essential to understanding the scientific program. If key demonstrations of equipment or techniques are to be presented during the tour, then visitors must remain together. The TL should be prepared to be flexible about this part of the agenda.

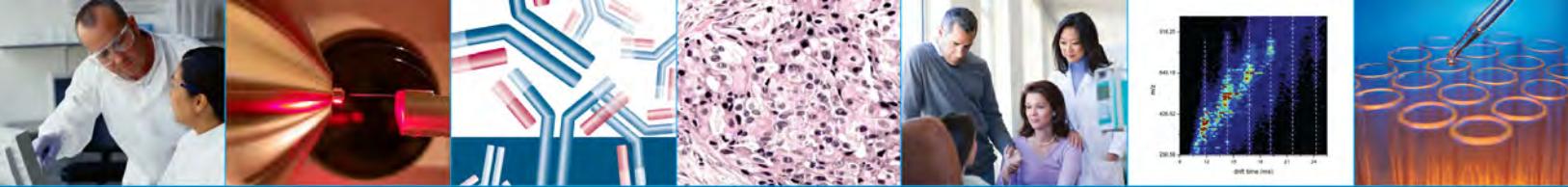
## **Budget Session**

A formal budget session should be scheduled after the scientific presentations to allow CPTC to make certain that they understand each request. (schedule only if deemed necessary)

## **Conclusion of the Site Visit**

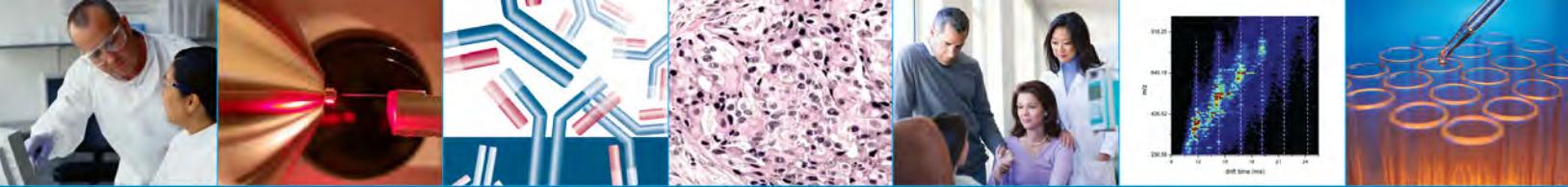
The site visit ends with a meeting with the TL and any key people s/he may designate. At this time, participants may present anything perceived to have not been covered adequately earlier, as well as answer any additional questions, especially those of a confidential nature.

\*Additional items, such as meeting with administrators (Deans, department chair, program administrator, etc.) may be made part of a site visit, but should be kept as short as possible.



## CPTAC Center Site Visit Evaluation Checklist

Criteria	Yes	No	Comments
<i>Proteomics Experiments</i>			
Interlab experiments			
- is the lab fully participating in Studies 1-7?			
Two variants of MS technologies			
- are the two variants actively being tested?			
Alternative to MS (affinity)			
- is the affinity capture platform yet being explored?			
Standard reference materials			
- are reference materials regularly used to calibrate instruments and check performance?			
Endpoints of experimental design			
- is the lab receiving input on design and analysis from expert statistical counsel?			
<i>Data</i>			
- is the lab contributing data to the CPTAC central storage site?			
- what software is being developed that will become a resource to the community?			
- is there a plan to make data and software caBIG™ compliant?			
- is there a data sharing plan?			
- is the data sharing plan being implemented?			
<i>Biospecimens</i>			
Biospecimen collection			
- how many patients have been accrued in the CPTAC study to date?			
Selection of candidates for platform evaluation			
- what candidates have been selected?			
Evaluate detection of cancer-relevant proteomic changes			
- are clinical samples being analyzed for reproducible features?			
<i>SOPs</i>			
- are sample prep SOPs followed in this lab?			
- are proteomic analysis SOPs followed in this lab?			
- are data analysis SOPs followed in this lab?			
Potential IP			
- Are there any IP issues NCI should be aware of?			



## 4.7 eProtein Newsletters

- I. Summer 2009
- II. Spring 2009
- III. Winter 2008



## CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

Advancing Protein Science for Personalized Medicine



# eProtein

## *Letter from the Director*



Dear Colleagues,

In June of this year, members of the Clinical Proteomic Technologies for Cancer (CPTC) community provided a program update to NCI's Board of Scientific Advisors (BSA), which was very well received. The BSA, consisting of 35 members from a number of disciplines in science and medicine, advises NCI's senior leadership on a wide variety of matters concerning scientific program policy and the progress and future direction of extramural research programs.

The presentations were especially focused on the more robust and efficient protein biomarker development pipeline that has been developed by this initiative. Simply said, CPTC is restructuring this pipeline to include a verification, or pre-validation step, which serves as a bridge between biomarker discovery and clinical validation. Verification may provide a very reliable GO/NO GO decision point and potentially save the medical diagnostic industry millions of dollars and many years of development because only the strongest candidates will move into clinical validation—and with much greater confidence.

In just a few short years, CPTC investigators have made significant advances in the field that will affect the way every investigator does protein biomarker discovery research. To learn about this endeavor, and other tremendous advances being made by this initiative, including community resources and data release policies, I encourage you to attend our upcoming annual meeting this October. I look forward to seeing you there! ■

**A Clinical Proteomic  
Technologies for Cancer  
initiative publication  
that builds connections  
throughout the  
proteomics community**

## *Under-Represented Students: Training the Next Generation of Cancer Research Scientists*

The National Institutes of Health Emerging Technologies Continuing Umbrella of Research Experiences (ET CURE) pilot program addresses the need for a diverse cancer research community in the 21st century to reflect the nation's ethnic heterogeneity. In support of ET CURE, NCI's Center to Reduce Cancer Health Disparities (CRCHD)

*continued on page 2*

## *An Advocate's Perspective A Response to "The Promise and the Reality of Proteomics" Webinar*

**Elda Railey**

*Co-Founder*

*Research Advocacy Network*

For years, we have been promised more personalized medicine and targeted therapies for cancer, and today we have learned more of the real promise for cancer detection and treatment through the study of proteomics. The issue of cost savings through early detection methods is very important as the strain on our

*continued on page 3*

## *In This Issue*

***In Vitro Diagnostic Tests  
for Cancer: Navigating  
the FDA Approval Process***

**4**

***Researcher Spotlight:  
John C. Chaput, Ph.D.  
Next Generation Affinity  
Reagents for Cancer  
Biomarker Detection***

**5**

***Researcher Spotlight:  
Amanda Paulovich, M.D., Ph.D.  
Targeted Proteomics:  
Relieving a Bottleneck in  
the Biomarker Pipeline***

**6**

***Industry News***

**7**

***Upcoming Events***

**8**

***Reagents Data Portal  
Newly Released  
Antigens and Antibodies***

**8**

## *Under-Represented Students: Training the Next Generation of Cancer Research Scientists*

(continued from cover)

has provided principal investigators with the opportunity to plan and implement a research training program in emerging technologies through research supplements for high school and undergraduate students from underserved populations.

The goals of the ET CURE initiative are to:

1. Create a pipeline of underserved students and investigators in the fields of emerging and advanced technologies;
2. Increase the number of scientists from underserved populations with training in the elective disciplines of focus, such as nanotechnology, clinical proteomics, bioinformatics, biophotonics, and cancer health disparities;
3. Enhance the application of emerging technologies to cancer research through increased training and educational opportunities; and
4. Foster academic, scientific, and multi-disciplinary research excellence to culminate the emergence of a mature investigator capable of securing competitive advanced research funding.

There are a number of domestic institutions involved in the ET CURE initiative. In this pilot program, LeeAnn Bailey, Ph.D., Program Director of CRCHD, will determine how each institution recruits students, the types of applicants they receive, the types of programs that have been set up, and what proves successful. The pilot will then be followed by a larger scale program.

Amanda Paulovich, M.D., Ph.D., of the Fred Hutchinson Cancer Research Center

*"I was given tremendous research opportunities as an undergraduate student at Carnegie Mellon University, and I want to give that back."*

(FHCRC) and a member of the Clinical Proteomic Technology Assessment for Cancer (CPTAC) Network, is an active participant in the ET CURE program. "My lab is involved in clinical proteomics, and our goal is to give budding students exposure to the field," explains Paulovich.

The original plan was to have a single undergraduate student from an underserved or minority population join the Paulovich lab to gain exposure to clinical proteomic technologies. This plan changed, however, when three applications in particular stood out.

"One is a biology major, one a chemistry/biology double major, and one a computer science major. This stuck out to me because in real projects in the lab that are funded on grants, we tend to have a three person team: a biochemist or biologist, a chemist, and a bioinformaticist," outlines Paulovich. "We thought, wouldn't it be fun to put these three together as a team? So that's what we did."

The first three students included Brianna Byers and Tim Nguyen, both from the University of Washington, and Christina Tieu from Pacific Lutheran University. Christina recently left FHCRC to continue a successful career in biomedicine through the M.D./Ph.D. program at the Mayo Clinic. A local high school student, Tao Large, from The Northwest School, has since joined the lab as her replacement.

The curriculum for these students includes four separate activities that are



*(From left to right) Tao Large, Tim Nguyen, and Brianna Byers*

geared towards preparing them for a successful career in biomedical research.

First, the students are working collaboratively on a project using state-of-the-art proteomic approaches to discover cellular responses to DNA damage. This project has been designed to cover topics spanning basic biology through advanced mass spectrometry and data analysis. Every member of the Paulovich lab is actively engaged in assisting the students in these experiments. "When they came into the lab, the students received a list of topics they would touch on during their experiments as well as a roster identifying the appropriate contact person in the lab that the students could talk to about each of the relevant topics," says Paulovich. "This prevented just one person from getting overwhelmed with responsibility while also ensuring that the students develop relationships with all of the lab members. Hence, the whole lab is taking part in the effort, helping the students navigate the experiments from beginning to end."

*continued on page 3*

(continued from previous page)

Second, the students are also developing their presentation skills. Paulovich has provided the students with a list of topics on which they will prepare presentations and deliver them to the entire lab group, which currently includes 25 members. "The students recently gave their first round of presentations, and they exceeded my expectations! It is very invigorating to have their energy and enthusiasm around the lab," says Paulovich.

Third, in order to give the students a real taste of what it means to be in biomedical research, they are gaining critical experience in the NIH grant application process. Following a crash course in this process (given by Paulovich), the students are completing a mock grant application that assumes they have to convince a panel of reviewers to fund their clinical proteomics project. "This exercise is useful not only for exposing them to the grant process but also for forcing them to think critically about their experiments and to learn relevant background information related to their project," notes Paulovich.

And finally, the students are taking part in weekly faculty seminar series and writing workshops that have been set up for all summer interns at FHCRC. This provides them with the opportunity to network with faculty members and other interns.

"The critical thing is to get them excited about the field," explains Paulovich. "I was given tremendous research opportunities as an undergraduate student at Carnegie Mellon University, and I want to give that back." ■

## An Advocate's Perspective

### A Response to "The Promise and the Reality of Proteomics" Webinar (continued from cover)

#### Research Advocacy Network

healthcare system is widely felt. However, the cost savings of early detection methods are negated when the results are not reliable, or the costs of the tests outweigh their validity.

Part of the "omics" promise is the development of better candidates for drug therapies and early cancer diagnostics for improved cure rates and reduced costs of treatment, both in human suffering and in dollars. This promise is especially enticing in harder to diagnose cancers such as squamous cell head and neck cancer.

I must admit that I was still confused about the difference between genetics/genomics and proteomics. After studying the materials from NCI, I have a better understanding that genes give a glimpse of what MAY occur, and proteomics can help understand what is happening in REAL TIME. The reality for a patient is that it is not important what type of technology or "omics" science results in the best detection methods and personalized treatment choices, but it is very important to patients that the results returned by these technologies are accurate and reliable.

When donating biospecimens for research, it matters that the "piece of me," whether it is blood, serum, or another biospecimen, is used to gain the maximum amount of information and contribute to the knowledge base to fight cancer. We also want to be assured that our privacy is protected.

Even though I have been in cancer advocacy for many years, it was not easy to make proteomics a concept relevant to our everyday lives. Yet this area is where much of the research investment dollars

and the state of science are headed. It was published recently that colon cancer patients who knew about targeted therapies were more likely to receive those treatments. We will probably find that patients who know about these important early detection methods are more likely to utilize them.

For advocates to be helpful we do need to understand what the proteomics pipeline really contains and what the outcomes of the work have been in the past. We need to understand what the barriers and promoters of the knowledge are along the way, and how researchers across disciplines are sharing these technologies and knowledge so that these methods can be integrated into clinical practice.

*"The reality for a patient is that it is not important what type of technology or "omics" science results in the best detection methods and personalized treatment choices, but it is very important to patients that the results returned by these technologies are accurate and reliable."*

We hope that there will be future opportunities to partner with advocates to truly fulfill the promise of proteomics and "team science." We also recommend that advocates representing the patient perspective be considered an integral part of the "team" to participate as a partner in prioritizing discussions and problem solving, to review educational materials, and to serve as a communications channel to disseminate information about the promise and realities of proteomics to our constituencies. ■

## In Vitro Diagnostic Tests for Cancer: Navigating the FDA Approval Process



**Jeffrey N. Gibbs, J.D.**  
*Director*  
*Hyman, Phelps & McNamara, PC*  
*Washington, D.C.*

The rapid growth in proteomics has provided new insights not only into biological processes but also into tools for creating novel *in vitro* diagnostic (IVD) assays. Researchers are already using proteomic-based assays to generate diagnostic information, and as the knowledge base and technologies advance, the rate of development of new IVDs will increase dramatically.

However, creating and clinically validating proteomic-based IVDs does not necessarily mean that they will be incorporated into medical practice. Discovering, developing, and validating a proteomic-based assay is necessary, but not sufficient, for an assay to be commercialized. There are still other hurdles, including compliance with regulatory requirements.

Under federal law, new IVDs must be reviewed by the Food and Drug Administration (FDA) before they can be sold in the United States. The FDA is the gatekeeper for all new diagnostic assays that are commercially distributed in the U.S. While the FDA has exempted some low-risk, well-understood assays from the need for prior review, proteomic-based assays will not qualify for this exemption.

*"Discovering, developing, and validating a proteomic-based assay is necessary, but not sufficient, for an assay to be commercialized."*

Developers of proteomic-based diagnostic test kits should expect that they will need to secure FDA marketing authorization before they can launch their kits in the U.S.

There is an alternative route to introducing the assay: offering it as a laboratory developed test. This pathway will be discovered in a separate article in the next issue of *eProtein*.

To complete the FDA review process, companies will need to negotiate multiple steps. Some of the key steps are summarized briefly below.

**1. Identify a specific intended use:** It is difficult to overstate the importance of developing a precisely worded intended use statement at an early stage. The regulatory process is heavily influenced by the intended use. For example, the FDA does not let companies sell a test labeled as "an X Cancer Assay." Rather, the intended use must be more specific and contain information about diagnosis, prognosis, production, screening, monitoring, etc. of a particular type of cancer. The intended use statement may also need to describe how the assay fits into the diagnostic paradigm (e.g., for use by primary care physicians as compared to oncologists). These seemingly subtle word choices can have a major impact on the regulatory process. Even if the wording changes later on, companies need to develop a working intended use statement at the outset.

**2. Developing a protocol:** Companies developing novel proteomic-based IVDs should expect that they will need to conduct a clinical study. The study must be consistent with the proposed intended use. The protocol should carefully address, among other issues, source

of specimens, clinical comparator (other FDA-cleared method, clinical diagnosis, etc.), and statistical methodology.

**3. Meeting with the FDA:** In general, there is no obligation to get permission from the FDA before beginning a diagnostic study. Nevertheless, obtaining the FDA's feedback can be extremely helpful, particularly for novel assays. Companies with new kinds of proteomic assays would generally be well served by meeting with the FDA before beginning a clinical validation study. Topics at the meeting could include proposed intended use, key elements of the draft protocol, the regulatory pathway, and statistics. Companies must prepare carefully for these "pre-IDE" (Investigational Device Exemption) meetings and listen carefully to FDA's comments.

**4. Conducting the study:** Because the clinical validation study is intended to support an FDA marketing application, it must meet FDA regulatory requirements, which may entail monitoring study sites and laboratories. Clinical trial agreements are essential; institutional review board (IRB) approval and informed consent may be necessary. The reliability and integrity of the data need to be established.

**5. Conducting preclinical studies:** Companies will also need to conduct a variety of preclinical studies to assess analytical performance. The FDA has developed guidelines relating to preclinical studies. While these are not legally binding, applicants should review the pertinent guidelines before beginning these studies.

**6. Submitting the FDA application:** There are two major routes for obtaining

*continued on page 7*

## Researcher Spotlight:

### Next Generation Affinity Reagents for Cancer Biomarker Detection



**John C. Chaput, Ph.D.**  
*Research Investigator*  
*The Biodesign Institute at*  
*Arizona State University*

The discovery that tumors leak proteins and peptides into bodily fluids has led to the idea that it may be possible to diagnose cancer at pre-symptomatic stages or access a patient's response to treatment by monitoring specific cancer biomarkers present in human blood and urine. While over 1,000 cancer biomarkers have been described in the literature, only a small fraction of these targets have been independently validated, and an even smaller fraction have a medical diagnostic available for their detection.

One reason for the limited number of validated cancer biomarkers is the lack of high quality affinity reagents needed to detect and bind these targets in complex biological mixtures. Many researchers consider this bottleneck to be a grand challenge in basic and applied biomedicine as the need for high quality affinity reagents is now impacting many large-scale projects that attempt to explore the nature and function of the human proteome. Overcoming this problem will likely

*"Overcoming this problem will likely require transformative ideas that shift the current paradigm..."*

require transformative ideas that shift the current paradigm away from methodologies that are costly and time consuming and focus instead on novel solutions that are capable of changing the way in which protein affinity reagents are created.

One approach to relieving the antibody bottleneck is to develop a chemical strategy for making protein affinity reagents that no longer relies on animal immunization or iterative rounds of *in vitro* selection and amplification as the primary means of discovery.

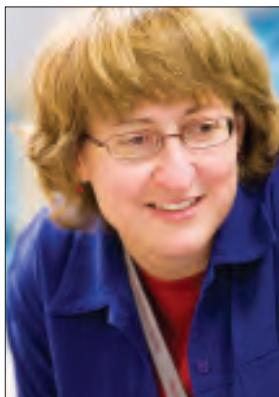
To address this problem, our lab is working to develop a versatile two-step strategy for creating artificial protein affinity reagents that we call *synbodies*. In the first step, non-competing peptide ligands are discovered by array-based or single-pass, high-throughput screening methods that bind to different sites on the surface of a desired protein target. In the second step, combinations of orthogonal peptides are screened on a synthetic DNA scaffold to identify the optimal peptide pair and peptide pair distance required to transform two modest affinity ligands into a single high affinity protein binding agent. This strategy, which is general and amenable to high-throughput, has the potential to become an enabling technology by providing a simple method for creating high quality synthetic antibodies to any water-soluble protein.

In a proof-of-principle demonstration, we generated synbodies to the yeast protein Gal80 and the human blood protein transferrin. In both cases, assembly of the peptides at optimal distances on the DNA scaffold resulted in a synthetic antibody with affinity and specificity similar to a typical antibody. We found that synbodies function in a standard enzyme-linked immunosorbent assay (ELISA) and in an immunoprecipitation assay, which suggests that simple chemical reagents represent a viable alternative to traditional antibodies.

In partnership with CPTC's Advanced Proteomics Platforms and Computational Sciences program, we are applying this technology to develop synthetic affinity reagents to important cancer biomarkers, such as growth factor receptor-bound protein 2 (Grb2), prostate-specific antigen (PSA), and cancer antigen 125 (CA-125). We are particularly interested in evaluating the potential use of synbodies relative to well-characterized monoclonal and polyclonal antibodies. The outcome of these experiments will help establish the general utility of synbodies as future protein affinity reagents for proteomics and cancer research. ■

## Researcher Spotlight:

### Targeted Proteomics: Relieving a Bottleneck in the Biomarker Pipeline



Amanda Paulovich, M.D., Ph.D.  
*Assistant Member*  
*Fred Hutchinson Cancer Research Center*

Discovering protein biomarker candidates is relatively easy. In fact, hundreds to a thousand protein biomarker candidates are typically discovered at one time using genomic or proteomic technologies such as mass spectrometry. The problem, however, is that most of these candidates are not clinically useful biomarkers. The true biomarkers must be culled from this lengthy list of candidates, which is a very time-consuming, costly, and inaccurate process.

This gap in the biomarker development pipeline—between discovery and clinical validation—results from the lack of available assays for testing candidate biomarkers. This limitation has proven to be a significant barrier for clinical proteomics, and it partly explains why most biomarker candidates never reach clinical testing. A faster and less expensive assay is needed in order to reduce the time and cost of evaluating novel potential cancer diagnostic biomarkers.

Amanda Paulovich, M.D., Ph.D., Assistant Member of the Fred Hutchinson Cancer Research Center and a member of the Clinical Proteomic Technology Assessment for Cancer (CPTAC) Network, is developing a verification technology that will help to determine which biomarker candidates are worth pursuing in the clinic. Specifically, her laboratory is collaborating closely with Drs. Leigh Anderson (Plasma Proteome Institute), Steve Carr (The Broad Institute of MIT and Harvard), Terry Pearson (University of Victoria), and Steve Skates (Massachusetts General Hospital) to test a targeted type of mass spectrometry (MS), called multiple reaction monitoring (MRM-MS), that can measure specific proteins in a highly multiplexed fashion. Combined with an enrichment technology, stable isotope standards and capture by anti-peptide antibodies (SISCAPA), MRM assays achieve sufficient sensitivity for measuring candidate protein biomarkers in the ng/ml or lower range in plasma.

"These assays are faster and cheaper to develop than conventional immunoassays, and the multiplex capability [ability to test multiple biomarkers in one test] is quite high," explains Paulovich. "We are able to test far larger numbers of candidates than has been possible in the past, presumably improving our odds of identifying clinically useful markers."

SISCAPA-MRM-MS may serve as the much needed bridge between biomarker discovery and clinical validation. Restructuring the biomarker

*"This assay technology is still at an early stage and is undergoing tremendous refinement to make it more acceptable for potential clinical applications, which is what our group is really focused on."*

development pipeline with the addition of this new assay technology will ensure that only the strongest biomarker candidates will move into clinical validation. "This assay technology is still at an early stage and is undergoing refinement to make it more acceptable for potential clinical applications, which is what our group is really focused on," says Paulovich.

Paulovich, along with Jeff Whiteaker, Ph.D., and Lei Zhao in her laboratory, recently participated in a collaborative effort within the CPTAC network, demonstrating the reproducibility of MRM-based assays across laboratories, which is a critical characteristic for detection of disease-specific biomarkers. This study was published in the July ■ 2009 issue of *Nature Biotechnology*.

*A Multi-site Assessment of Precision and Reproducibility of Multiple Reaction Monitoring-based Measurements By the NCI-CPTAC Network: Toward Quantitative Protein Biomarker Verification in Human Plasma. Nat. Biotechnol. [Epub 2009 Jun 28]*

## Industry News

### *Study Supports New Bridge Technology for Biomarker Development Pipeline: MRM-MS*

A team of CPTAC researchers has demonstrated that a new method for detecting and quantifying protein biomarkers in body fluids, multiple reaction monitoring mass spectrometry (MRM-MS), may ultimately make it possible to screen multiple biomarkers in hundreds of patient samples, thus ensuring that only the strongest biomarker candidates will advance down the development pipeline. The goal of this research is to reduce the time and cost of developing cancer diagnostic tests, ultimately increasing the number of such tests in the clinic so cancer can be caught at its earliest stages.

"These findings are significant because they provide a potential solution for eliminating one of the major hurdles in validating protein biomarkers for clinical use. Thousands of cancer biomarkers are discovered every day, but only a handful ever makes it through clinical validation. This is a critical roadblock because biomarkers have the potential to allow doctors to detect cancer in the earliest stages, when treatment provides the greatest chances of survival," says John E. Niederhuber, M.D., NCI director. "The critical limiting factor to date in validating biomarkers for clinical use has been the lack of standardized technologies and methodologies in the biomarker discovery and validation process, and this research may solve that dilemma."

The study results were published in the online version of *Nature Biotechnology* on June 28, 2009. [Click here](#) to read the full press release. [Click here](#) to read coverage in the NCI Cancer Bulletin.

### *imaGenes to Distribute Highly Characterized Monoclonal Antibodies Produced by CPTC*

imaGenes ([www.imagenes-bio.de](http://www.imagenes-bio.de)), a premier provider of genome research services in Europe, will distribute monoclonal antibodies created and characterized by CPTC. CPTC's antibody characterization program, a component of its Proteomic Reagents and Resources core, uses standard operating procedures to create highly characterized monoclonal antibodies to human proteins associated with cancer for research use (<http://antibodies.cancer.gov>). "This resource will accelerate biomarker discovery and validation and will ultimately assist to rapidly advance the use of new biomarkers in clinical practice," says Johannes Maurer, imaGenes' Director of Genomic Products & Marketing.

### *Advancing Principles for Data Sharing by Proteomics Researchers*

Leaders in proteomics are pushing to develop a set of principles to guide data sharing in this field. A [Journal of Proteome Research paper](#), which resulted from the 2008 International Summit on Proteomics Data Release and Sharing Policy, held in Amsterdam, outlines the challenges facing such efforts.

Read the [full story](#) in the *Journal of Proteome Research*. No subscription is required.

### *In Vitro Diagnostic Tests for Cancer: Navigating the FDA Approval Process (continued from page 4)*

FDA marketing authorization. The 510(k) premarket notification requires the applicant to demonstrate that its assay is "substantially equivalent" to a legally marketed "predicate device," i.e., a device cleared by the FDA or marketed before May 28, 1976. (The latter option is unlikely here.) The key is finding another 510(k)-cleared device with the same or roughly similar intended use. The other primary route is the premarket approval application (PMA). PMAs are more complicated than 510(k)s and also are subject to more controls and regulatory requirements once on the market. In general, the 510(k) route will be

preferred. Both kinds of applications require careful preparation and attention to detail. A third alternative is the "*de novo* automatic classification," but that has been used in only a very small percentage of applications.

Once the FDA receives a 510(k), they can approve it (technically called a "clearance"), ask questions, or reject it (found "not substantially equivalent"). The FDA will respond to the 510(k) within ninety days of submission. If they ask for more information, the ninety-day clock may reset upon receipt of the company's reply. For PMAs, the outcomes are similar: approval, ask questions, or disapproval. The review clock for PMAs is 180 days.

Getting clearance or approval does not discharge all regulatory obligations. Once an IVD is on the market, companies must comply with multiple FDA post-marketing regulatory requirements.

Proteomic-based technologies offer exciting opportunities for improving clinical diagnoses. However, before these tests can be offered commercially as kits, they will need to successfully navigate the FDA review process. Just as the development of the assay itself requires careful planning and scientific rigor, so does the FDA process. ■



## CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

Advancing Protein Science for Personalized Medicine



## Upcoming Events

**September 26-30, 2009**

*HUPO VIII World Congress*  
The Westin Harbour Castle  
Toronto, Canada

**October 5-7, 2009**

*Clinical Proteomic  
Technologies for Cancer  
Annual Meeting: Advancing  
Protein Science for  
Personalized medicine*  
Hyatt Regency Bethesda  
Bethesda, Md

**October 7-9, 2009**

*Innovative Molecular Analysis  
Technologies (IMAT) Program Meeting*  
Organized by: NCI  
Hyatt Regency Bethesda  
Bethesda, Md

**November 5-6, 2009**

*Translating Novel Biomarkers to  
Clinical Practice: Role and Opportunities  
for the Clinical Laboratory*  
Hosted by: American Association  
for Clinical Chemistry (AACC)  
The Marriott Bethesda Hotel  
& Conference Center  
Bethesda, Md

For a full list of upcoming events, visit

<http://proteomics.cancer.gov/mediacenter/events>.

## Contact Information

For more information about the CPTC, please visit  
<http://proteomics.cancer.gov>, or contact us at:

National Cancer Institute  
Office of Technology & Industrial Relations  
ATTN: Clinical Proteomic Technologies for Cancer  
31 Center Drive, MSC 2580  
Bethesda, Md 20892-2580  
Email: [cancer.proteomics@mail.nih.gov](mailto:cancer.proteomics@mail.nih.gov)

**The NCI Clinical Proteomic Technologies for Cancer initiative seeks to foster the building of an integrated foundation of proteomic technologies, data, reagents and reference materials, and analysis systems to systematically advance the application of protein science to accelerate discovery and clinical research in cancer.**

## Reagents Data Portal

<http://antibodies.cancer.gov>

<http://dshb.biology.uiowa.edu>

*Newly Released Antigens and Antibodies*

Antigen	Antibody
14-3-3 Sigma	CPTC-SFN-1 CPTC-SFN-2 CPTC-SFN-3
BCL2-like 2	CPTC-BCL2L2-1 CPTC-BCL2L2-2 CPTC-BCL2L2-3
Calcyclin	CPTC-Calcyclin-1 CPTC-Calcyclin-2
Chloride Intracellular Channel 1	CPTC-CLIC1-1 CPTC-CLIC1-2
Fascin	CPTC-Fascin-1 CPTC-Fascin-2 CPTC-Fascin-3
Glutathione S Transferase M1	CPTC-GST M1-5 CPTC-GST M1-6 CPTC-GST M1-7
Melanoma Antigen Family A, 4	CPTC-MAGEA4-1 CPTC-MAGEA4-2 CPTC-MAGEA4-3
MethylCpG Binding Protein 1	CPTC-MBD1-1 CPTC-MBD1-2 CPTC-MBD1-3
Protein Phosphatase 2A	CPTC-PP2A-1 CPTC-PP2A-2 CPTC-PP2A-3 CPTC-PP2A-4
Ubiquitin conjugating enzyme E2C	CPTC-UBE2C-1



## CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

Advancing Protein Science for Personalized Medicine



# eProtein

## *Letter from the Director*



Dear Colleagues,

We are extremely pleased with the amount of interest the Clinical Proteomic Technologies for Cancer (CPTC) initiative has received at various national forums over the past few months. At US HUPO this past February, for example, our Clinical Proteomic Technology Assessment for Cancer (CPTAC) centers presented their latest research findings to a packed, standing-room only audience. In addition, at the American Association of Cancer Research (AACR) 2009 Annual Meeting, CPTAC successfully hosted a special Methods Workshop Session titled "CPTC Proteomics Technology Platforms for the Cancer Biomarker Pipeline." It has become increasingly apparent that the scientific community supports our efforts to build more refined, efficient, and reliable biomarker discovery and verification pipelines. These pipelines are anticipated to produce better credentialed candidate leads, ultimately accelerating the discovery of new cancer biomarkers. As more of our findings go into the public domain, we look forward to enhancing our interactions with the greater scientific and clinical communities. ■

**A Clinical Proteomic  
Technologies for Cancer  
initiative publication  
that builds connections  
throughout the  
proteomics community**

## *NCI and FDA Engaged in Active Dialogue on the Analytical Validation of Clinical Proteomic Technologies*

The National Cancer Institute (NCI)-U.S. Food and Drug Administration (FDA) Interagency Oncology Task Force (IOTF) Molecular Diagnostics subcommittee held a workshop in Cambridge, Mass. on October 30, 2008, bringing together almost 60 participants representing the

*continued on page 2*

## *A Physician's Perspective Protein Markers of Risk and Prognosis: A Role for Proteomics*

While breast cancer remains the most common cancer of North American women, population-based statistics indicate that for the first time in history, incidence and mortality rates have been falling over the past 10 years. Reduced mortality rates are likely the consequence of earlier diagnosis and more effective treatment modalities developed over the past few decades. One is tempted to speculate that even earlier diagnosis will result in even

*continued on page 3*

## *In This Issue*

**Lack of Metrics and Standards  
in Proteomic Discovery  
Technologies Hinders  
Innovative Research for  
Rare Diseases**

4

**The Patient as Stakeholder:  
CPTC Engagement with  
Patient Advocates**

5

**Researcher Spotlight:  
Daniel Liebler, Ph.D.  
Analyzing the Subtleties  
of Cancer-Relevant  
"Proteotypes"**

6

**Researcher Spotlight:  
Richard Smith, Ph.D.  
Pushing the Limits  
of Protein Biomarker  
Candidate Detection**

7

**Industry News**

8

**Upcoming Events**

9

**Reagents Data Portal**

9

## NCI and FDA Engaged in Active Dialogue on the Analytical Validation of Clinical Proteomic Technologies

(continued from cover)

National Institutes of Health (NIH), FDA, industry, academia, and standards organizations. These key stakeholders in the proteomics community gathered to explore the regulatory requirements that will be needed to validate protein-based marker panels and any new technologies (hardware) for their intended use.

According to Mansfield, holding the workshop in conjunction with the CPTC annual meeting was a perfect opportunity to involve CPTC scientists who are currently working through the issues that the FDA will need to address when reviewing 510(k) submissions for proteomic technologies such as mass spectrometry and affinity arrays.

*"There's really no guidance for multiplex proteomic assays. We've seen some mass spectrometry assays come through here, but none of them have been for proteomics ... and there are unique issues when you start to do a multiple test in a single tube or platform."*

The IOTF was formed in 2003 to enhance and accelerate the overall process of developing new cancer diagnostics and therapeutics. "The molecular diagnostics group was formed a couple of years ago, and Henry [Rodriguez] and I were named as the subcommittee co-chairs. Together, we decided we would focus on an area of unmet need, which is proteomics," explains Elizabeth Mansfield, Ph.D., a Senior Policy Analyst in the Office of In Vitro Diagnostic Devices at the FDA.

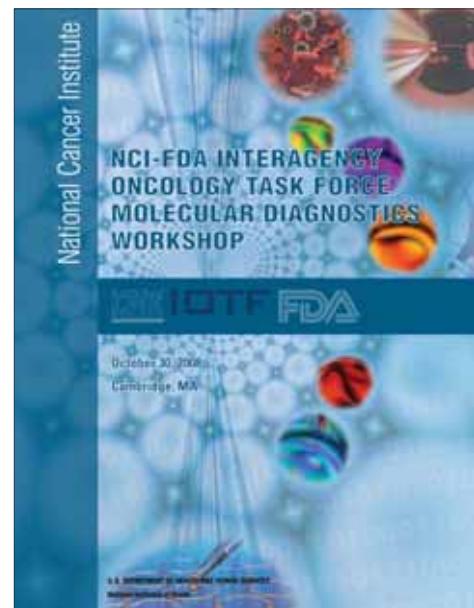
"There's really no guidance for multiplex proteomic assays. We've seen some mass spectrometry assays come through here, but none of them have been for proteomics; they have all been metabolites, so we're kind of naïve on that [the protein] side," says Mansfield. "In terms of multiplex immuno-affinity assays, we actually don't see very many of those either, and there are unique issues when you start to do a multiplex test in a single tube or platform."

There were six case studies given by representatives from the FDA and other members of the proteomics community that addressed issues expected to be faced in the 510(k) submission process when the technologies are ready, including how to qualify a proteomic technology, specimen and population issues, statistical issues, and understanding the regulatory pathway to commercialization. The result was a highly engaging workshop with both groups—the FDA and the proteomics community—posing relevant questions to each other with the goal of understanding the challenges and needs of each side.

"We laid the groundwork for a good understanding of each other," says Mansfield. However, the interaction will not end there. Invited participants are in the process of developing two sets of documents that will keep the conversation—and the momentum—going. First, the group of invited

participants will develop a workshop summary for publication that will discuss the analytical validation issues that specific proteomic technologies should address when seeking FDA approval. Second, the group will create mock 510(k) regulatory submissions for two technologies—mass spectrometry and affinity platforms—drawing on information gathered during the workshop and extending the dynamic conversations held throughout the day.

Together, these documents will help orient the FDA to proteomic technologies in novel diagnostics and serve as a springboard for guidance to the proteomics community. ■



Information on the NCI-FDA IOTF Molecular Diagnostics Workshop deliverables will be posted at <http://proteomics.cancer.gov> as it becomes available.

## A Physician's Perspective

### Protein Markers of Risk and Prognosis: A Role for Proteomics (continued from cover)



**Gabriel N. Hortobagyi, M.D., F.A.C.P.**

*Professor of Medicine*

*Nellie B. Connally Chair in Breast Cancer*

*Chairman, Department of Breast*

*Medical Oncology*

*Director, Multidisciplinary Breast Cancer Research Program*

*Department of Breast Medical Oncology*

*The University of Texas M D. Anderson*

*Cancer Center*

*Houston, Texas, USA*

better outcomes, hence the enthusiasm with which high-throughput methods for genomics, proteomics, and metabolomics have been greeted.

Gene expression profiles have been explored over the past decade, and their clinical application revealed moderately accurate prognostic value and for some available assays, the ability to predict response or benefit from specific systemic therapies. However, gene expression profiles are obtained from tumor material, so they do not lend themselves for prediction of risk (except for carriers of known adverse mutations, e.g., BRCA1/2 carriers), or for early diagnosis. Furthermore, while gene profiling provides an overview of the genome, it does not always reveal how genomic changes can lead to disease.

It was for this reason that experiments to identify specific protein profiles that can be identified in the bloodstream have been pursued with vigor in recent years. The

underlying hypothesis was that tumor cells, exhibiting a number of molecular genetic anomalies, would produce abnormal types or quantities of proteins, that these proteins could be identified in biological fluids *in vivo*, and that they would lead to the identification of subjects at risk and could also be used to effect diagnosis of breast cancer at an earlier time than currently available diagnostics would permit. Furthermore, such markers might enhance the diagnostic accuracy of mammography and sonography and might serve to enrich populations of healthy subjects who would most benefit from systematic screening. When this hypothesis first surfaced, the available technology was inadequate to test it.

Despite improvements in technology over the past few years, there remains a general lack of understanding of the reproducibility and sensitivity of various discovery proteomics platforms for the detection of real differences between samples. Lack of knowledge has fueled skepticism in the scientific and medical communities about the capabilities of proteomics to be able to deliver. Clearly, there are a multitude of analytical and preanalytical variables that pose significant challenges

to the translation of proteomic discoveries into clinical applications. The best way forward to developing this knowledge is through highly collaborative and integrated consortia of expert proteomic groups such as those formed through the NCI's CPTAC program. Properly designed inter-laboratory studies applied to relevant, carefully annotated biological samples, with the results analyzed with appropriate informatics and biostatistics will be essential to forming a clear understanding of the capabilities and limitations of proteomic technologies.

We must combine the necessary knowledge, expertise, and technology to develop these novel markers; however, the challenges involved in this process are daunting. In addition to conceptual and technological progress, our methodology must be flawless to identify and validate markers and to determine in prospective trials their clinical value and contribution to the management of diagnosis, treatment, and prevention of breast cancer. Addressing and reducing all layers of variability at every step of the biomarker pipeline through multidisciplinary approaches will be vital to realizing the promise of proteomics.■

## Honors

Dr. Hortobagyi's research includes combination chemotherapy regimens, presurgical chemotherapy, and targeted therapies for all stages of breast cancer. He has contributed more than 800 articles to scientific journals, authored and co-authored 13 books, and contributed over 130 chapters to textbooks. For his efforts in breast cancer research, Dr. Hortobagyi has received worldwide honors. In 2001, President Jacques Chirac named him Chevalier of the Order of la Legion d'Honneur de France. In 2003, Dr. Hortobagyi received the Glen Robbins Award in Breast Cancer Research from the New York Cancer Society and the Metropolitan Breast Cancer Group, and the Bristol-Myers Squibb 2003 Horizon Scientific Award. The Mexican Society of Oncology named him the 2005 World Leader in Oncology. Dr. Hortobagyi was elected President of the American Society of Clinical Oncology (ASCO) for the 2006–2007 term.

## Lack of Metrics and Standards in Proteomic Discovery Technologies Hinders Innovative Research for Rare Diseases



Jeffrey Kaufman

The personal experiences of Marnie Kaufman and her husband Jeff with adenoid cystic carcinoma (ACC) gave rise to the Adenoid Cystic Carcinoma Research Foundation (ACCRF). Mrs. Kaufman is a survivor of ACC—a rare, slow-growing cancer of the head and neck that typically originates in the salivary glands. Although the removal of her parotid gland and a course of radiation treatment sent her cancer into remission, the slow and persistent growth of ACC necessitates regular screenings for metastasis.

Dismayed by a lack of ongoing research in ACC, the Kaufmans decided to battle the disease on another front. In December 2005, they founded the ACCRF. Mr. Kaufman would eventually leave his role as a senior portfolio manager with Putnam Investors to take on the full-time job of Executive Director of the ACCRF.

The mission of the ACCRF moves away from a traditional advocacy focus on patient support and education; the Foundation concentrates instead on coordinating ongoing research efforts to accelerate improvements in cancer treatment and the discovery of a cure. The approach of the ACCRF is simple:

advance discovery by identifying scientists and institutions working on the best research platforms and approach them with proposals for research partnerships. In doing so, the ACCRF seeks to establish a virtual network of ACC-focused researchers and clinicians to share resources and research findings.

During its short history, the ACCRF has developed a broad portfolio of research investments based on a wide array of platforms. Partnerships with the Sanger Institute, Göteborg University in Sweden, Harvard Medical School, and Johns Hopkins University have researchers developing a central biobank of ACC specimens, exploring gene mutations associated with ACC, and identifying gene targets through RNA interference techniques.

Despite the innovative and cutting-edge nature of the research funded by the ACCRF, its scientific advisory board has expressed hesitance in supporting proteomics-based research. "Due to concerns about the ability to replicate data collected through proteomics research, our organization has elected to focus our research investments in other areas," says Mr. Kaufman.

The ACCRF is not alone in recognizing a lack of standards in proteomics research methods. The NCI attributes the vast discrepancy between the number of cancer protein biomarkers that have been

described in scientific literature to date—over 1,200—and the surprisingly few that have transitioned into clinical applications to a lack of standardized procedures.

While ACCRF has not formally partnered with CPTC thus far, they are very encouraged by the work that the CPTC researchers are doing. "CPTC is providing a valuable service in addressing the challenge of validating the output of proteomics research," Mr. Kaufman states. "I am confident that the work of CPTC will have an important impact on the future of proteomics research initiatives of the ACCRF." ■

*"Due to concerns about the ability to replicate data collected through proteomics research, our organization has elected to focus our research investments in other areas."*

## The Patient as Stakeholder: CPTC Engagement with Patient Advocates



Elizabeth Neilson

The ultimate goal of the CPTC initiative is to advance protein biomarkers into molecular diagnostic tests for the early detection and treatment of cancer, ensuring that patients reap the rewards of the molecular revolution faster and more effectively. While optimizing clinical proteomic technologies is crucial to this goal, CPTC recognizes that their work cannot be completed without considering another perspective—the patients eagerly awaiting these clinical breakthroughs.

To gain insight into the patient perspective, CPTC established a connection with NCI's Consumer Advocates in Research and Related Activities (CARRA) organization. Since 2007, CARRA advocates have played an increasingly vital role in developing and executing the CPTC mission and objectives.

CARRA advocates offer a rich perspective on the experience of a cancer patient. Volunteers with the organization have backgrounds as cancer survivors, caregivers, or they have at least two years of involvement in cancer-related activities. Additionally, they must also volunteer with a cancer advocacy or support organization. "The volunteers' personal connections with cancer, combined with exposure to others' experiences with the disease through their volunteer activities, ensures that CARRA members are well

*"By responding to the feedback provided during our consumer testing, CPTC was able to produce an end product that clearly communicates the potential of proteomics research to patients."*

equipped to represent the needs of the broader cancer community," says Elizabeth Neilson, an Advocacy Relations Manager in NCI's Office of Advocacy Relations.

The brochure, *Clinical Cancer Proteomics: What It Means and What It Means for You*, provides a tangible example of the two-way information link between CPTC and CARRA. The five-page brochure is a means to communicate the complexities of proteomics research to the general public. While CPTC was responsible for the development of the brochure content, CARRA provided valuable feedback from advocates that impacted the final product.

By "testing" the brochure with cancer survivors and caregivers with a wide range of backgrounds—from attorneys, to business owners, to college professors—CARRA discovered that an initial draft of the brochure was too technical for the general public. "By responding to the feedback provided during our consumer testing, CPTC was able to produce an end product that clearly communicates the potential of proteomics research to patients," Ms. Neilson comments. "The brochure also provides a much-needed resource for the [CPTC] leadership and researchers to explain their research to the general public."

CPTC and CARRA continue to facilitate communication between the research and the patient. A webinar designed to educate patient advocates about proteomics research was conducted on March 19, (see *Industry News*, page 8) and a targeted research effort with advocates involved

with NCI-funded Specialized Program of Research Excellence (SPORE) programs and Cooperative Groups is in the pipeline to identify gaps in the understanding of proteomics research.

The partnership between CPTC and CARRA is a truly symbiotic relationship. Patients and advocates are informed about research in proteomics-based technologies with the potential to improve detection and treatment of cancer, and this newly informed group ultimately helps accelerate the adoption of these technologies as they become clinically available. Meanwhile, researchers and clinicians have a constant reminder of the patients that their work will impact. As this information sharing continues, the collaboration between CPTC and CARRA strengthens, and a better outcome for patients comes within reach. ■



## Researcher Spotlight: Daniel Liebler, Ph.D.

### Analyzing the Subtleties of Cancer-Relevant "Proteotypes"



Daniel Liebler, Ph.D.

Proteins are involved in almost all biological activities, including disease, and researchers are trying to identify proteins that can be used as biomarkers to better detect and treat cancer. The proteome—the collection of proteins in a cell or tissue—provides a rich source of biological information for these challenging studies.

Daniel Liebler, Ph.D., of Vanderbilt University, and current Chair of the CPTAC Program Coordinating Committee, is trying to compare proteomes between cancer-relevant tissues to identify biomarkers. "The term we are using to represent the proteomic differences [between tissues] is proteotypes, which is analogous to the term genotype," explains Liebler, who introduced this concept during a presentation at the NIH on February 6, 2009 as part of the Protein Interest Group seminar series.

"Shotgun proteomics" is the most powerful platform that researchers use for analyzing complex proteomes. Analogous to the shotgun genome sequencing strategy that was used to sequence the human genome, proteins are "cut up" into little pieces called peptides, data is collected on the peptides using analytical

platforms, and then computers put the pieces back together to deduce what proteins are present.

Although powerful, variable performance of this approach can prevent detection of very subtle differences between proteomes. To overcome this barrier, Liebler and colleagues developed and implemented a new analytical platform for shotgun proteomics that they are currently applying to colon cancer samples.

"A 2mg piece of tissues gives us approximately 4,000 protein identifications. If we compare [the proteotypes] of a normal colon tissue and a large adenoma, the difference between these data sets is in the neighborhood of 150 proteins," explains Liebler, who pointed out that cancer and normal tissue contain two very different proteotypes.

Can we compare proteotypes that contain much more subtle differences? For example, how many proteins will differ between two stage 2 colon cancers, one of which is likely to recur as a more aggressive disease in a few years after surgery and the other, which is likely to never recur? "What is the difference between these much more subtle phenotypes? They're both cancers. They both look the same to the pathologist, and in fact, there aren't any good markers that allow us to predict which tumor's going to be the aggressive recurring disease," explains Liebler.

To determine if their platform is sensitive enough to detect such subtle differences, Liebler conducted proof of concept studies on cell lines because they are very well characterized. They analyzed the proteomes from a collection of colon cancer cell lines that differ only

in expression of DNA mismatch repair genes. They found that although the proteomes are essentially identical, the DNA mismatch repair-deficient cell lines differ in the expression of a number of proteins whose genes are known to be affected by this pathway. Thus, these proteotypes reflect known biology.

Following the success of these proof of concept studies, Liebler and colleagues are now using their shotgun proteomics platform to analyze tissues retrospectively from patients with stage 2 colon cancers, with goal of characterizing distinct proteotypes for colon cancer recurrence. If successful, their work could help pave the way for the next generation of cancer diagnostic testing—proteotyping.

*In December 2008, Dr. Liebler was elected as a fellow of the American Association for the Advancement of Science (AAAS). ▀*

*"The term we are using to represent the proteomic differences [between tissues] is proteotypes, which is analogous to the term genotype."*

## Researcher Spotlight: Richard Smith, Ph.D.

### Pushing the Limits of Protein Biomarker Candidate Detection



**Richard Smith, Ph.D.**

A biomarker is a molecule, such as a protein, that is present in the body's tissues or fluids and can alert doctors to disease. Finding valid biomarkers is of high interest, for example, for the early detection and more effective treatment of cancer. It is anticipated that cancer cells "leak" proteins into bodily fluids, suggesting that clinicians can potentially detect protein biomarkers with a simple blood or urine test. However, discovering these biomarkers is not so simple.

Proteins exist in a wide range of concentrations in human plasma, spanning several orders of magnitude, and the proteins of interest—the ones most likely to be the telltale signs of cancer—are expected to be present in extremely tiny quantities. Identifying and measuring these dilute proteins has proven to be a significant challenge using current technologies.

To address this problem, Richard Smith, Ph.D., Director of the NIH Research Resource for Integrative Proteomics at the Pacific Northwest National Laboratory, is developing a significantly more sensitive and robust cancer biomarker assessment platform for the analysis of human blood plasma. This platform,

which is being developed through CPTC's Advanced Proteomic Platforms and Computational Sciences program, aims to make biomarker discovery and the subsequent follow up (verification studies) much more effective.

"We've developed a new platform that's based on a combination of ion mobility separations that follow fast liquid chromatography separations and ultra-fast mass spectrometry," says Smith. In the past year alone, Smith and his team have shown that this platform has the capability to expand the dynamic range of measurements significantly so that low abundance proteins—at least one order of magnitude lower in abundance than can be detected in plasma using current technologies—can be detected. And do so faster.

"This new technology opens up a large number of potential biomarkers that have previously been beyond the range of what we can do," says Smith. There is also good indication that they can potentially add another order of magnitude onto their current detection levels, which would significantly increase sensitivity. Three patents have been received for parts of this new platform's technology.

Efforts in biomarker discovery generally involve identifying many proteins in a very limited number of samples followed by verification studies in a much larger number of samples, using more targeted and sensitive measurements as a means to whittle down what is often a large set of candidates prior to clinical validation. This platform will allow researchers to take many more protein measurements during the initial discovery phase, and it has the potential to help find which candidates can be useful biomarkers for cancer detection. By effectively collapsing

the discovery and verification phases due to its higher measurement throughput and sensitivity, this technology may eliminate the need for, or greatly shorten, that second step.

"With the success of the platform and its utilization in biomarker discovery, the biomarker discovery process will be faster and much more effective," explains Smith. "What we're trying to do is make a considerable leap in the performance of measurement technology so as to make biomarker discovery both faster and more effective." ■

*"With the success of the platform and its utilization in biomarker discovery, the biomarker discovery process will be faster and much more effective. What we're trying to do is make a considerable leap in the performance of measurement technology so as to make biomarker discovery both faster and more effective."*

## Industry News

### Molecular & Cellular Proteomics Introduces New Guidelines for Clinical Proteomics Manuscripts

In keeping with the tradition that it established when instituting standards for articles reporting protein identifications,<sup>1</sup> *Molecular & Cellular Proteomics (MCP)*, published by the American Society for Biochemistry and Molecular Biology, has now developed guidelines for manuscripts that report clinically relevant proteomic studies.<sup>2</sup> The guidelines were initially drafted at a meeting that was held in Copenhagen in April of 2008 with the participation of about two dozen stakeholders covering various areas of expertise, such as pathology, biobanking, medical law and ethics, statistics, various proteomic technologies, as well as different clinical areas. Following a period of public consultation, the guidelines were duly edited and adopted in January of this year. The guidelines describe information that is required before *MCP* will consider any germane submission further as well as discretionary information that, if available, will be helpful to reviewers and will add to the value of the reported data. The editors and staff of *MCP* view these guidelines as an essential step in the process of ensuring that the rapidly developing area of clinical proteomics advances in a productive and effective fashion. The Guidelines may be viewed at [www.mcponline.org](http://www.mcponline.org).

<sup>1</sup> R. A. Bradshaw, A. L. Burlingame, S. Carr, and R. Aebersold (2006) Reporting Protein Identification Data: The Next Generation of Guidelines. *Mol. Cell. Proteomics*, 5: 787 - 788.

<sup>2</sup> J. E. Celis, S. A. Carr, and R. A. Bradshaw (2008) New Guidelines for Clinical Proteomics Manuscripts. *Mol. Cell. Proteomics*.

### *hPDQ: A Near-Term Practical Step towards a Large-Scale Human Proteome Project*

A Human Proteome Project (HPP), a natural successor to the Human Genome Project, will one day bring enormous benefit to patients in the form of personalized medicine. Although scientific objectives and technical merits to such an endeavor remain unclear, a group of researchers involving academia and industry have proposed an alternative, tactical approach for providing a more basic map of the human proteome. Instead of mapping every human protein, the Human Proteome Detection and Quantitation (hPDQ) initiative takes a targeted approach that would enable researchers to measure defined collections of human proteins in biological samples with high sensitivity and absolute specificity. A two-year pilot would seek to target 2,000 human proteins that have biomarker potential.<sup>1</sup> This would be followed by an additional 18,500 proteins over a five year period. This initiative, which can be thought of as an incremental first step towards a massive, large-scale effort, could be just what HPP will need to get off the ground and running.

<sup>1</sup> N. L. Anderson, N. G. Anderson, T. W. Pearson, C. H. Borchers, A. G. Paulovich, S. D. Patterson, M. Gillette, R. Aebersold, and S. A. Carr (2009) A Human Proteome Detection and Quantitation Project: hPDQ. *Mol. Cell. Proteomics*, Jan 7 [Epub ahead of print]

### *Promise and Reality of Proteomics Webinar*

The NCI Office of Advocacy Relations hosted a webinar on March 19, 2009 on the role of protein science in the early detection of cancer as part of the Understanding NCI: Toll-Free Teleconference Series. Dr. Rodriguez and other speakers discussed the challenges facing clinical proteomics and the innovative ways that NCI's CPTC initiative aims to develop new, more refined, efficient, and reliable biomarker discovery and verification pipelines. This webinar can be found at: <http://proteomics.cancer.gov/library/webinars.asp>

Speakers include:

- Henry Rodriguez, Ph.D., M.B.A., Director, Clinical Proteomic Technologies for Cancer, NCI Center of Strategic Scientific Initiatives
- Amanda G. Paulovich, M.D., Ph.D., Oncologist and Cancer Geneticist, Director, Early Detection Initiative, Fred Hutchinson Cancer Research Center
- Elda Railey, Co-Founder of Research Advocacy Network

### *Podcast Series*

"**The Promise of Proteomics for Personalized Medicine**" is the first of a three part podcast series, made available on March 20, 2009, developed to provide listeners with information about research advances being made in proteomics. This podcast can be found at: <http://proteomics.cancer.gov/library/podcasts.asp>.



## CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

Advancing Protein Science for Personalized Medicine

### Upcoming Events

**May 12-13, 2009**

Data Release Workshop

Organized by: Genome Canada

Ottawa, Canada

**October 5-7, 2009**

Clinical Proteomic Technologies for Cancer

3<sup>rd</sup> Annual Meeting Hyatt Regency Bethesda

Bethesda, MD

For a full list of upcoming events, visit

<http://proteomics.cancer.gov/mediacenter/events>.

*We are deeply saddened to hear about the passing of our colleague, Dr. Katheryn Resing, from the University of Colorado at Boulder. We extend our heartfelt condolences to her family, friends, and colleagues during this difficult time.*

### Contact Information

For more information about the CPTC, please visit

<http://proteomics.cancer.gov>, or contact us at:

National Cancer Institute  
 Office of Technology & Industrial Relations  
 ATTN: Clinical Proteomic Technologies for Cancer  
 31 Center Drive, MSC 2580  
 Bethesda, MD 20892-2580  
 Email: [cancer.proteomics@mail.nih.gov](mailto:cancer.proteomics@mail.nih.gov)

**The NCI Clinical Proteomic Technologies for Cancer initiative seeks to foster the building of an integrated foundation of proteomic technologies, data, reagents and reference materials, and analysis systems to systematically advance the application of protein science to accelerate discovery and clinical research in cancer.**



### Reagents Data Portal

<http://antibodies.cancer.gov>

<http://dshb.biology.uiowa.edu>

Newly Released Antigens and Antibodies

Antigen	Antibody
Aldo keto reductase Family 1 Member B1	CPTC-AKR1B1-1 CPTC-AKR1B1-2 CPTC-AKR1B1-3
Annexin A1	CPTC-ANXA1-1 CPTC-ANXA1-2 CPTC-ANXA1-3
Chromogranin A	CPTC-CHGA-1 CPTC-CHGA-2 CPTC-CHGA-3
Crystallin Alpha B	CPTC-CRYAB-1 CPTC-CRYAB-2 CPTC-CRYAB-3
Ezrin (p 81)	CPTC-Ezrin-1
Gelsolin	CPTC-Gelsolin-1
Glutamate Cysteine Ligase Regulatory Subunit	CPTC-GCLM-1 CPTC-GCLM-2 CPTC-GCLM-3
Glutathione S Transferase M1	CPTC-M1-1 CPTC-M1-2 CPTC-M1-3 CPTC-M1-4
Glutathione S Transferase M2	CPTC-M2-1 CPTC-M2-2 CPTC-M2-3
Interleukin 18	CPTC-IL18-1 CPTC-IL18-3
Lactoglutathione Lyase	CPTC-GLO-1 CPTC-GLO-2 CPTC-GLO-3
Ornithine Decarboxylase 1	CPTC-ODC1-1 CPTC-ODC1-2 CPTC-ODC1-3
Peroxiredoxin 4	CPTC-PRDX4-1 CPTC-PRDX4-2 CPTC-PRDX4-3
Squamous Cell Carcinoma Antigen 1	CPTC-SERPINB3-1 CPTC-SERPINB3-2 CPTC-SERPINB3-3



# eProtein

## *Letter from the Director*



Dear Colleagues,

Since the Clinical Proteomic Technologies for Cancer (CPTC) initiative was launched just over two years ago, tremendous progress has been made towards reducing the layers of systems variability that plague candidate protein biomarker discovery. This is not surprising given that we have an amazing scientific network—the best minds in proteomics—representing nearly 50 federal, academic, and private sector organizations who are deeply committed to open and collaborative science for the sake of the entire cancer community. A great deal of work still remains before us, but we are confident that the investment that the National Cancer Institute (NCI) is making in proteomics today will pave the road for clinical translation tomorrow.

CPTC is pleased to launch this online quarterly newsletter, *eProtein*, as a way for the proteomics community to keep abreast of the many exciting achievements being made within the CPTC programs as well as by individual colleagues all across the nation. We are committed to ensuring the success of this important initiative because together we are building the foundation for clinical cancer proteomics. ■

## *The 2<sup>nd</sup> Annual CPTC Meeting, Cambridge, Mass.*

CPTC held its second annual meeting in Cambridge, Mass. on October 28–29, 2008, bringing together more than 200 participants representing the full gamut of scientific fields that contribute to the initiative's mission to review the technological progress made over the previous year.

Giving a sense of the links between CPTC and other technology focused initiatives supported by NCI, the

*continued on page 2*

**A Clinical Proteomic Technologies for Cancer initiative publication that builds connections throughout the proteomics community**

## In This Issue

**Facilitating Data Sharing and Release in Proteomics** 4

**SBIR Program Helps Integrate Technology Development Efforts by CPTC and Small Businesses** 5

**Researcher Spotlight: David Tabb Making Sense of the Complexity with User-Friendly Tools** 6

**Researcher Spotlight: Xiaolian Gao Visions of High-Throughput Clinical Proteomics** 7

**Upcoming Events** 8

**Reagents Data Portal Newly Released Antigens and Antibodies** 8

**In the Next Issue** 8

*continued on page 3*

## The 2<sup>nd</sup> Annual CPTC Meeting, Cambridge, Mass. (continued from cover)

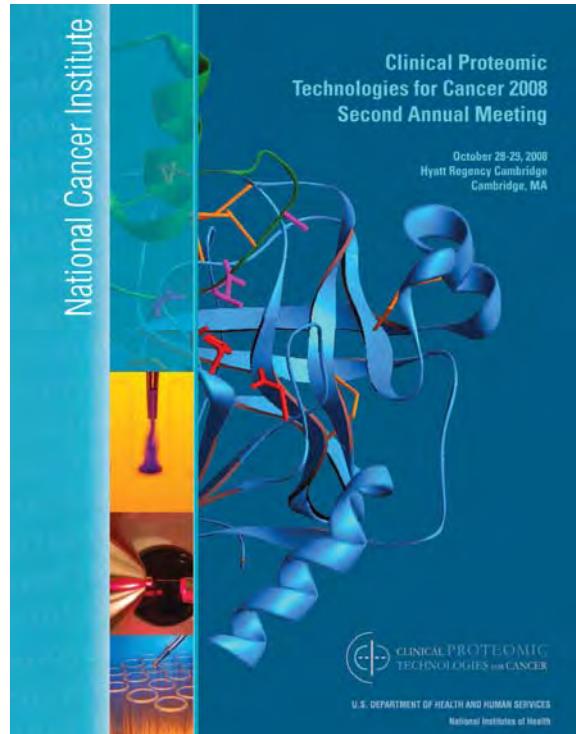
first day of the meeting was held jointly with members of NCI's Innovative Molecular Analysis Technologies (IMAT) program.<sup>1</sup> Several talks featured technologies and techniques developed by IMAT-supported investigators that have subsequently been applied to projects supported by CPTC, highlighting the importance of integrated technology development in cancer proteomics research in particular and in cancer research in general. The meeting also included talks and posters featuring research conducted through CPTC's three components: the Clinical Proteomic Technology Assessment for Cancer (CPTAC) program, Advanced Proteomic Platforms and Computational Sciences, and the Proteomic Reagents and Resources Core (see *New Antibody Portal Puts Well-Characterized Antibodies at Researchers' Fingertips*, cover and page 3). Both days featured keynote addresses by researchers speaking on their experiences in integrated research. David Altshuler, M.D., Ph.D., a founding

member of the Eli M. and Edythe L. Broad Institute of MIT and Harvard and Director of the Institute's Program in Medical and Population Genetics, spoke of the lessons learned from conducting large-scale genomics research and how those lessons could apply to large-scale proteomics. In particular, he noted ways of avoiding pitfalls in validating variations, such as early development of robust, comprehensive, and scalable tools for determining systematic associations—an issue that, he noted, CPTC is well on its way to addressing by focusing on technology development up front. Altshuler concluded his remarks by reminding attendees of the importance of grounding new discoveries in human biology before jumping to conclusions about their importance.

The keynote address on the second day, given by Vamsi Mootha, M.D., of the Broad Institute and Massachusetts General Hospital, focused on integrative genomic, proteomic, and metabolomic

research on mitochondrial diseases. The mitochondrial proteome has not yet been fully defined, but it may contain between 1,200 and 1,500 proteins, only 13 of which have been associated with genes found in the mitochondrial genome; the rest are encoded by nuclear genes. Mootha's talk outlined his work to develop a mitochondrial protein catalog, called MitoCarta, which currently contains 1,098 mitochondrial proteins. With this information in hand, he has started probing the ancestry of numerous mitochondrial proteins and applying that knowledge clinically to explore rare familial diseases caused by breakdowns in mitochondrial respiration oxidative phosphorylation.

In his closing remarks, CPTC Director Henry Rodriguez, Ph.D., M.B.A., noted that the initiative had produced some very good outputs since its launch two years ago. Rodriguez also mentioned that while there had been a learning curve associated with the initiative, they had shown that team-based science could be



*"CPTC is definitely on track to meeting the goal of optimizing cancer proteomics research."*

— Moyez Dharsee  
Director, Informatics  
Ontario Cancer Biomarker Network

very successful, and that the steps that had been undertaken thus far had laid the groundwork for CPTC's future success.

*The third CPTC Annual Meeting will be held on October 5-7, 2009, in Bethesda, Md. Information on the meeting will be posted at <http://proteomics.cancer.gov> as it becomes available.* ■

<sup>1</sup> To learn more about IMAT, visit the program Web site at <http://imat.cancer.gov>.

## New Antibody Portal Puts Well-Characterized Antibodies at Researchers' Fingertips

(continued from cover)

appropriate for their experimental platform is available for their studies.

At the CPTC Annual Meeting in October 2008, the initiative's Proteomic Reagents and Resources Core announced the launch of the Reagents Data Portal (<http://cpti.abcc.ncifcrf.gov/>), a Web-based service open to the scientific community that helps scientists search for and access antibodies from a collection of highly characterized mouse monoclonal antibodies for cancer-associated proteins. According to Tara Hiltke, Ph.D., a Program Manager at CPTC, "Users can perform a keyword or alphabetical search to look up a protein, see the antibodies available for it, see the characterization information for those antibodies, choose the one that best fits their needs, and seamlessly order it from the repository at the University of Iowa's Developmental Studies Hybridoma Bank [DSHB]."

Antibodies in the collection are being generated against 1,261 tumor-associated proteins listed by Anderson and Polanski in 2006 as part of a collaboration between CPTC and several laboratories and companies.<sup>1</sup> Argonne National Laboratory (ANL) clones and expresses each of the proteins, which are then provided to private sector partners contracted for antibody generation through requests for proposals. Each contractor receives 40 proteins from ANL and generates 10 monoclonal IgG antibodies for each, ultimately submitting three of the 10 for in-depth characterization at four collaborating centers: NCI-Frederick, NCI's Center for Cancer Research Tissue Array Research Program (Bethesda, Md.), the Harvard Institute of Proteomics (Cambridge, Mass.), and the Human Protein Atlas at KTH Royal Institute of

*"We're trying to make as much information available to the community as possible on each antibody so that members can reproduce, trust, and use these antibodies."*

— Tara Hiltke, Ph.D.  
CPTC Program Manager

Technology (Stockholm, Sweden).

Each antibody is analyzed using:

- ELISA
- Immunohistochemistry
- Immuno-mass spectrometry
- Nucleic acid programmable protein arrays
- SDS-PAGE
- Surface plasmon resonance
- Tissue arrays
- Western blot

All of the antibodies and their associated hybridomas are deposited in the DSHB; once characterization is complete, they are made available to the public at a significantly discounted price.

A key advantage of the Portal is the depth of data available. All of the associated characterization information is available for each antibody, as is detailed information on each protein (i.e., sequence, molecular

weight, isoelectric point, alternate names, accession numbers, DNA source, and expression system). In addition, detailed standard operating procedures are posted for protein and antibody generation and characterization analyses. "We're trying to make as much information available to the community as possible on each antibody so that members can reproduce, trust, and use these antibodies, and so that they can serve a wide variety of applications," said Hiltke. ■

<sup>1</sup> Polanski M and Anderson NL. A list of candidate cancer biomarkers for targeted proteomics. *Biomark Insights*. 2006;2:1-48.



## Facilitating Data Sharing and Release in Proteomics

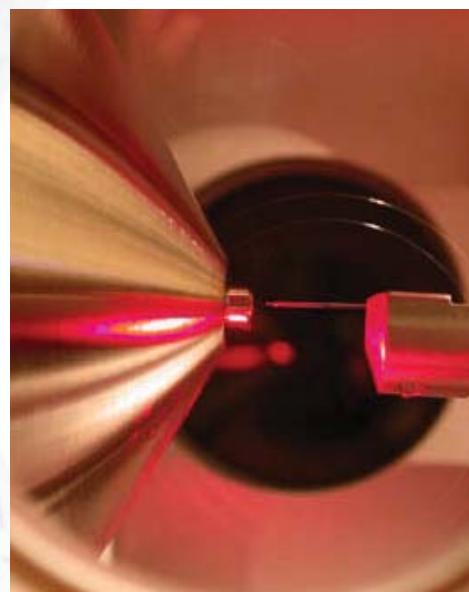
[Ed. The following is a condensed version of an editorial written for the *Journal of Proteome Research*.<sup>1</sup>]

Data sharing is standard practice among members of the genomics community, based on principles developed at a 1996 gathering in Bermuda and ultimately endorsed by all major parties in the Human Genome Project.<sup>2</sup> The widespread sharing of prepublication sequence data greatly accelerated the pace of genomic discovery. However, similar policies do not exist for proteomics research, a state of affairs currently seen as a significant obstacle to progress in the field.

in community resource projects should be required to release data once they are produced. Investigators working on individual projects, on the other hand, should release data upon publication in a peer-reviewed journal.

**What types of data should be released, and what kinds of metrics should be used to define data quality?** Participants agreed that high-quality, well-annotated raw data (for mass spectrometry and protein/affinity array data) would be

encourage rich annotation, and develop seamless submission procedures.



*Their task: to begin defining policies and practices that would govern and facilitate the release of proteomic data into the public domain along the Bermuda model.*

In August 2008, the NCI sponsored a summit in Amsterdam for members of the international proteomics community, including representatives from funding agencies, journals, and academic research centers. Their task: to begin defining policies and practices that would govern and facilitate the release of proteomic data into the public domain along the Bermuda model. The summit's participants addressed the following questions:

### When should data be released?

Participants agreed that the timing of data release should be governed by the type of project. Investigators taking part

the most reliable interchange format for data repositories. Metadata, information on data quality, and identification quality control will all be critical as well. Accessing these data would require development of the proper infrastructure (i.e., community supported standardized formats, controlled vocabularies and ontologies, minimal reporting requirements, and publicly available online repositories). Central repositories should develop their own thresholds for data quality metrics, in a coordinated manner with users and one another, to ensure interoperability.

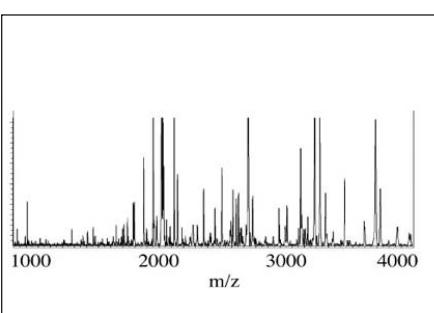
To fuel progress in proteomics research, data sharing cannot be voluntary; rather, it is up to scientists, journals, and funding agencies to take the necessary steps to ensure that all parties adhere to the standards for data release, ideally within a framework of tripartite responsibility akin to that created for genomics research.<sup>3</sup> Central repositories, for their part, should clearly define minimum submission requirements,

A white paper based on the discussions of the summit is forthcoming. ■

<sup>1</sup> Rodriguez H. International summit on proteomics data release and sharing policy. *J Proteome Res.* 2008 Oct 7. [Epub ahead of print]

<sup>2</sup> Policies on Release of Human Genomic Sequence Data. US Department of Energy, Human Genome Project. [www.ornl.gov/sci/techresources/Human\\_Genome/research/bermuda.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml). Accessed 22 October, 2008.

<sup>3</sup> Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. The Wellcome Trust. [http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy\\_communications/documents/web\\_document/wtd003207.pdf](http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf). January 2003.



## *SBIR Program Helps Integrate Technology Development Efforts by CPTC and Small Businesses*

Ready access to high-quality, standardized reagents is of great importance if the proteomics community is to catalyze biomarker discovery for reducing the burden of cancer. Numerous small businesses design and develop proteomic technologies for the accurate and powerful measurement of proteins and other biomolecules related to disease. Without well-characterized reagents, however, it is impossible to translate such platforms into products and services that could be used effectively by the cancer community.

One of CPTC's three component programs, the Proteomic Reagents and Resources Core, is tasked with providing the cancer community with the tools necessary to overcome technological and methodological barriers to developing and providing such reagents. To maximize the Core's capabilities and impact, CPTC is partnering with the biotechnology industry via the NCI's Small Business Innovation Research (SBIR) Program, a contract mechanism that supports early-stage research and development by small businesses. Through the SBIR program, CPTC aims to integrate its efforts with those of the biotechnology industry by encouraging and enabling companies developing proteomic technologies and platforms to adopt standardized, well-characterized reagents—including high-quality proteins and validated capture reagents (e.g., antibodies)—in the commercialization of new tools and kits for the cancer community.

CPTC has already awarded contracts based on SBIR requests for proposals released in 2007 and 2008 on such topics as "Development of Clinical Automated Multiplex Affinity Capture Technology for Detecting Low Abundance Cancer-related Proteins/Peptides" and "Advances in Protein Expression of Post-Translationally Modified Cancer Related Proteins." For instance, SBIR-awardee Rules-Based

Medicine Inc. is customizing CPTC-developed reagents for a quantitative, automated, Luminex-based 50-plexed immunoassay for the rapid detection of low abundance cancer-related proteins.

For fiscal year 2009, CPTC sought proposals on "Novel Antibody Epitope

Mapping Technologies," "Development of Novel Protein Expression Technologies for Glycosylated Cancer Related Proteins," and "Peptide Aptamers: New Tools to Capture and Study Protein Interactions in Lieu of Immunological Reagents." Contract awards for these topics are anticipated to be announced in the summer of 2009. ■

*Through the SBIR program, CPTC aims to integrate its efforts with those of the biotechnology industry.*

### **2007 SBIR Contract Recipients**

#### **Development of Clinical Automated Multiplex Affinity Capture Technology for Detecting Low Abundance Cancer-related Proteins/Peptides**

Meso Scale Diagnostics	Automated Multi-Array Platform for Cancer Biomarkers
Sequenom, Inc.	Sensitive Protein Detection Combining Mass Spectrometry
Quadraspec, Inc.	Highest Sensitivity Cancer Marker Array on Quadraspec's Bio-CD Platform
Rules-Based Medicine, Inc.	Automated Multiplexed Immunoassays for Rapid Quantification of Low Abundance Cancer-Related Proteins

#### **Development of Alternative Affinity Capture Reagents for Cancer Proteomics Research**

Allele Biotechnology & Pharmaceuticals	Yeast Single Chain Antibodies as Capture Reagents
Accacia International, Inc.	High-Throughput of Aptamers against Cancer Biomarkers

### **2008 SBIR Contract Recipients**

#### **Advances in Protein Expression of Post-Translationally Modified Cancer Related Proteins**

Rana Biosciences, Inc.	A Cell-Free System for High Yield Phosphoprotein Synthesis
------------------------	--

#### **Development of Clinical Quantitative Multiplex High-Throughput Mass Spectrometric Immunoassay for Detecting Low Abundance Cancer Related Proteins/Peptides in Bodily Fluids**

Intrinsic Bioprobes, Inc.	Multiplex Mass Spectrometric Immunoassays
Predictive Physiology and Medicine, Inc.	Immunoaffinity Capture Coupled with Ion Mobility

## *Researcher Spotlight: David Tabb, Ph.D.*

### *Making Sense of the Complexity with User-Friendly Tools*



David Tabb, Ph.D.

A bioinformatics researcher at Vanderbilt University, David Tabb, Ph.D., views his field from an integrated point of view, based on an underlying philosophy that software tools should be designed both for high performance and for comprehensibility by end-users.

Though he began his career by focusing his efforts on characterizing peptide fragmentation, Tabb now devotes his work to improving peptide identification through both database search and sequence tagging. He aims to integrate the two approaches to maximize the biological information produced from proteomics experiments.

Tabb initially joined CPTC as part of its Advanced Proteomic Platforms and Computational Sciences program. "CPTC was looking for algorithms to make sense of proteomic data, the area I find most interesting." He has since chaired the CPTAC Bioinformatics Working Group. "Funding from CPTC enables my team to develop research tools, while the Vanderbilt CPTAC funding enables us to apply and refine them for others," he says.

At all times, Tabb tries to keep research utility at the forefront of his design and development processes. "Too often, bioinformaticists view data only as data," he notes, "without looking at the biological picture or at the real need of laboratory biologists to understand how an algorithm produces a result." To that end, Tabb and his team work closely with the Vanderbilt CPTAC team, taking laboratory researchers' needs and perspectives into account. "My goal is to validate the tools my team and I are developing across multiple platforms and laboratories, so they can have the broadest biological utility."

For instance, because CPTAC's inter-laboratory experiments produce huge data sets that can be analyzed in multiple ways, the Tabb laboratory has emphasized algorithms that scale well for complex experiments instead of one small experiment at a time. Also, they try to address questions of usability. "Most tools are designed with a command line interface," Tabb says, "which few biologists are comfortable with." Thus, Tabb's group is working to add more user-friendly graphical interfaces to their algorithms. He is also building in connections to other software tools. The Tabb group is broadening the utility of IDPicker—their program for assembling protein sequences from raw peptide database identifications—by adding connections to spectrum review and sequence coverage tools and by incorporating the ability to

***Starting in the next issue, the Researcher Spotlight section will highlight investigators from the CPTAC network as well as the Advanced Proteomic Platforms and Computational Sciences program.***

accept identifications from any database search tool that can output pepXML-formatted data.

With the CPTAC Bioinformatics Working Group, Tabb is also working on methods to make CPTAC data available through the cancer Biomedical Informatics Grid (caBIG®), an NCI initiative intended to develop a nationwide information technology infrastructure for cancer research. "CPTAC program data will be made available to the scientific community via caBIG®," Tabb explains. "This presents its own challenges, as there are few tools or standards available in caBIG® for proteomics. We are working hard, though, to ensure that raw data will be available to anyone who wants them." ■

*"My goal is to validate the tools my team and I are developing across multiple platforms and laboratories, so they can have the broadest biological utility."*

— David Tabb, Ph.D.  
Assistant Professor of Biomedical Informatics  
Vanderbilt University School of Medicine

## Researcher Spotlight: Xiaolian Gao, Ph.D.

### Visions of High-Throughput Clinical Proteomics



Xiaolian Gao, Ph.D.

Small, inexpensive, fast, and able to assess the activity, structure, and interactions of thousands of genes at a time, DNA microarrays are a mainstay of genomics research. Protein microarrays have the potential to bring the same high-throughput analytical power to proteomics. "The microarray format is very familiar in the genomics world," says the University of Houston's Xiaolian Gao, Ph.D., "but it is still new in proteomics. However, the community is quickly recognizing its value as a means of running high-throughput experiments in ways that keep false positive rates low." Gao is contributing to the format's value by bringing array production and analysis together in the form of a microfabricated, high-density, addressable microarray platform being developed through CPTC's Advanced Proteomic Platforms and Computational Sciences program.

Gao's microarray technology is based on *in situ* peptide synthesis using a combination of microfluidics and methods akin to the photolithography techniques used to construct computer microchips. With these techniques, she can reproducibly achieve far higher peptide densities than most chips—3,000 to 4,000 peptides per

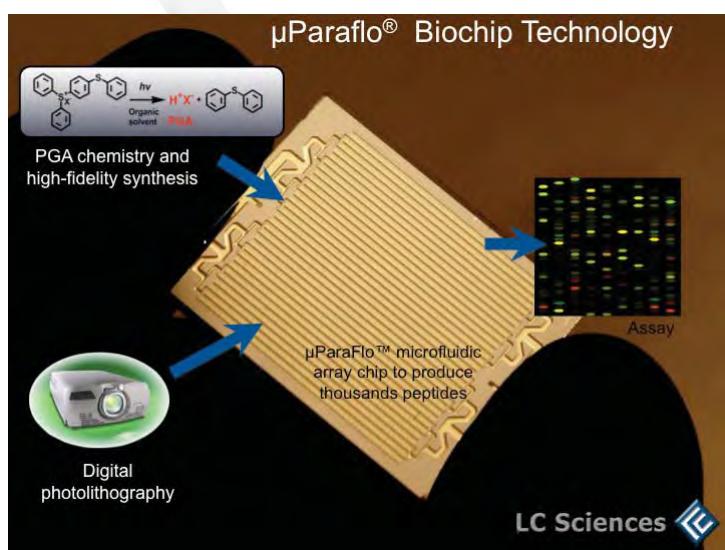
1-cm<sup>2</sup> array, as opposed to a few hundred peptides—using only microliter amounts of reagents per array.

The chips, which can be analyzed using standard DNA microarray fluorescence scanners, are also addressable. "The solution flow through our microfluidic chip is very much like chromatography," she says. "By controlling flow rate and temperature, we can run parallel synthesis and assays on a single chip. We can place individual peptide sequences in replicates and control and reference peptides at regular locations to minimize false positive measurements and increase stringency." Because of its flexibility, this array technology could bring new resolution to such experimental techniques as epitope mapping. "Generally, if you have a new antibody and want to find its binding site, you design and synthesize an epitope chip with 8- to 10-mer peptides derived from the antibody's target protein, add the antibody, and look for where it binds," said Gao. "With addressable *in situ* synthesis, we can both identify an antibody's core epitope sequence at a single amino acid resolution and quantify its binding constant. One microarray chip can contain as many

titrations as 41 96-microwell plates, allowing us to compare the specific binding of several antibodies simultaneously."

Gao is using the chips to study cancer-related signal transduction cascades by probing the interactions of phosphoprotein-binding proteins. "We can synthesize chips with peptide substrates that only recognize particular phosphoprotein-binding domains and develop signaling protein expression profiles for biospecimens such as serum or lysed cells. [The University of Minnesota's] Tongbin Li, Ph.D., and I are developing a database of these interactions that we hope will serve as the basis for peptide chip applications and computational systems biology efforts."

Gao envisions translating her microarray platform into an inexpensive, high-throughput system for clinical proteomics. "Ideally, a clinical laboratory would have a microfluidic chip analysis station, with which they could routinely take almost any patient specimen—blood, plasma, serum, digested tissue, maybe even urine—and look for diagnostic biomarkers or protein profiles." ■





# CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER



## Upcoming Events

February 22-25, 2009

US HUPO

5<sup>th</sup> Annual Conference

San Diego, CA

April 18-22, 2009

American Association of Cancer Research (AACR)

100<sup>th</sup> Annual Meeting

Denver, CO

October 5-7, 2009

Clinical Proteomic Technologies for Cancer

3<sup>rd</sup> Annual Meeting

Hyatt Regency Bethesda

Bethesda, MD

For a full list of upcoming events, visit

<http://proteomics.cancer.gov/mediacenter/events>.

## Reagents Data Portal

*Newly Released Antigens and Antibodies*

Antigen	Antibody
Interleukin 18	CPTC-IL18-2
Metastasin 100 calcium-binding protein A4 (Calvasculin)	CPTC-S100A4-1 CPTC-S100A4-2 CPTC-S100A4-3
Nucleoside Diphosphate Kinase B	CPTC-NME2-1 CPTC-NME2-2 CPTC-NME2-3
Ras-related C3 botulinum toxin substrate 1 (rho family small GTP binding protein Rac1)	CPTC-RAC1-1 CPTC-RAC1-2

## In the Next Issue of *eProtein*:

**A Success Story:** The Collaboration between CPTC and NCI's Center to Reduce Health Disparities for the Development of Training Opportunities in Emerging Technologies – Clinical Proteomics

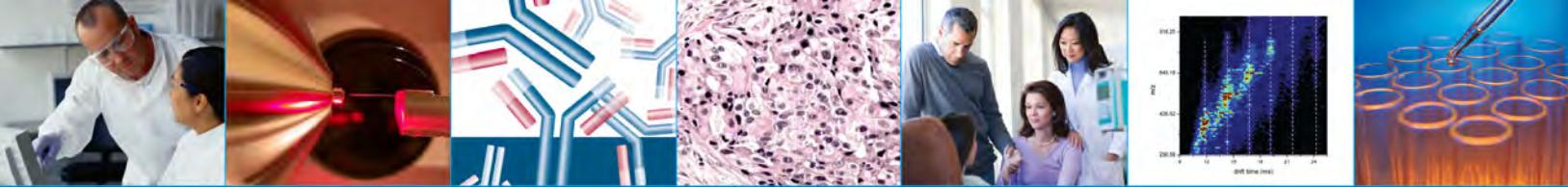
**Community Outreach:** How Engagement with Patient Advocates Enriches the CPTC Program

## Contact Information

For more information about the CTPC, please visit <http://proteomics.cancer.gov>, or contact us at:

National Cancer Institute  
Office of Technology &  
Industrial Relations  
ATTN: Clinical Proteomic Technologies  
for Cancer  
31 Center Drive, MSC 2580  
Bethesda, MD 20892-2580  
Email: [cancer.proteomics@mail.nih.gov](mailto:cancer.proteomics@mail.nih.gov)

**The NCI Clinical Proteomic Technologies for Cancer initiative seeks to foster the building of an integrated foundation of proteomic technologies, data, reagents and reference materials, and analysis systems to systematically advance the application of protein science to accelerate discovery and clinical research in cancer.**



## 4.8 Annual Investigator's Meeting Agendas

### 4.8.1 2007 Annual Meeting Agenda



National Cancer Institute  
National Institutes of Health  
U.S. Department of Health and Human Services

### NCI Clinical Proteomic Technologies for Cancer 2007 Annual Meeting

*October 24-25, 2007*

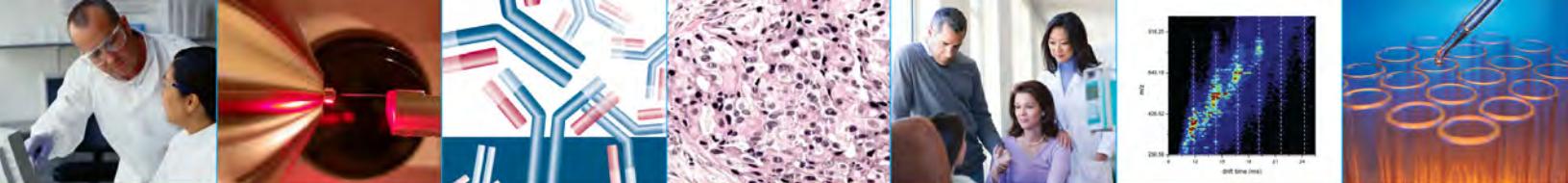
*Hilton Hotel  
Rockville, Maryland*

### AGENDA

---

#### ***Day 1, October 24***

7:30 a.m. - 8:30 a.m.	<b>Registration</b>
8:30 a.m. - 9:15 a.m.	<b>Welcome and Keynote Address</b>
8:30 a.m. - 8:35 a.m.	<b>Welcome</b> Henry Rodriguez, Ph.D., M.B.A. Director, Clinical Proteomic Technologies for Cancer National Cancer Institute, NIH
8:35 a.m. - 8:40 a.m.	<b>Perspective from a Cancer Patient Advocate</b> Ann McNeil, R.N. NCI Consumer Advocates in Research and Related Activities (CARRA) Member University of Miami
8:40 a.m. - 8:50 a.m.	<b>Special Address: Future Role of Proteomics in Molecular Oncology</b> John E. Niederhuber, M.D. Director National Cancer Institute, NIH



8:50 a.m. - 9:00 a.m.

**Overcoming Barriers in Molecular Diagnostics**

Anna D. Barker, Ph.D.

Deputy Director

National Cancer Institute, NIH

9:00 a.m. - 9:15 a.m.

**Overview of Clinical Proteomic Technologies for Cancer**

Henry Rodriguez, Ph.D., M.B.A.

Director, Clinical Proteomic Technologies for Cancer

National Cancer Institute, NIH

9:15 a.m. - 3:00 p.m.

**Clinical Proteomic Technology Assessment for Cancer (CPTAC) Program**

(Panel Session)

Moderator: Steven A. Carr, Ph.D.

Chair, CPTAC Program Coordinating Committee

Director, Proteomics

Broad Institute of MIT and Harvard

**Inter-Group Studies**

9:15 a.m. - 9:25 a.m.

**Introduction**

Steven A. Carr, Ph.D.

Director, Proteomics

Broad Institute of MIT and Harvard

9:25 a.m. - 10:05 a.m.

**Unbiased Discovery Platform Optimization: No Matrix**

Stephen E. Stein, Ph.D.

Director of Mass Spectrometry Data Center

National Institute of Standards and Technology

**Yeast Matrix**

Amanda Paulovich, M.D., Ph.D.

Assistant Member

Fred Hutchinson Cancer Research Center

10:05 a.m. - 10:25 a.m.

**Verification Platform Optimization Studies: MRM – Plasma Spiked With Proteins**

Steve Hall, Ph.D.

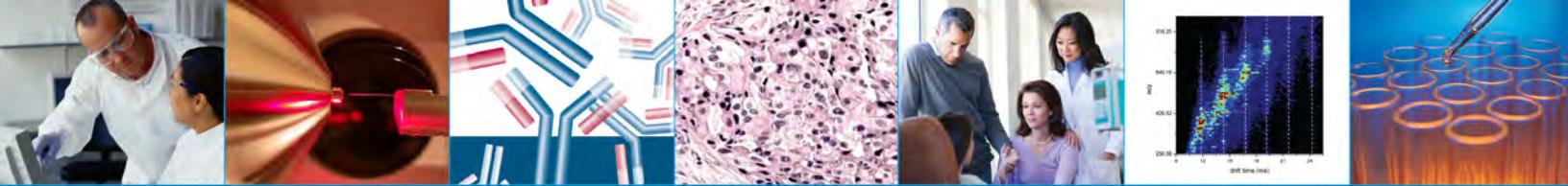
Associate Professor

University of California, San Francisco

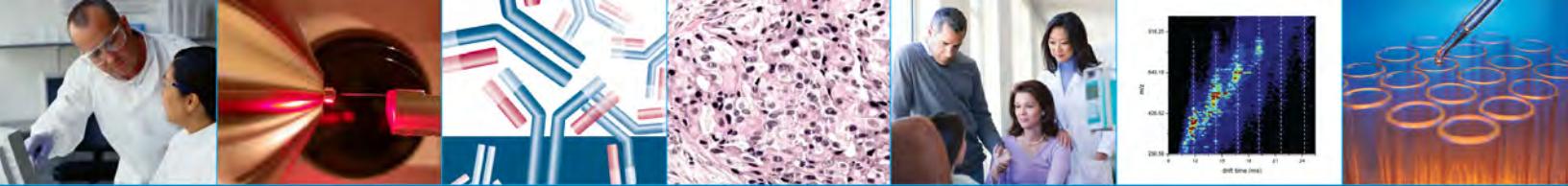
Steve Skates, Ph.D.

Assistant Professor in Biostatistics (Medicine)

Massachusetts General Hospital



10:25 a.m. - 10:40 a.m.	<b>Morning Break and Poster Session</b>
10:40 a.m. - 11:00 a.m.	<b>CPTAC Team Leader Panel Discussion (Inter-Group Studies)</b>
<b><i>Intra-Group Studies</i></b>	
11:00 a.m. - 11:30 a.m.	<b>Assessment of Serum Peptide Profiling to Detect Cancer-Specific Patterns</b> Paul Tempst, Ph.D. Professor Molecular Biology Program Memorial Sloan-Kettering Cancer Center
11:30 a.m. - 12 noon	<b>Targeted and Global Proteomic Strategies for Early Breast Cancer Detection</b> Susan J. Fisher, Ph.D. Professor University of California, San Francisco
12 noon - 1:30 p.m.	<b>Lunch and Official Poster Session</b>
1:30 p.m. - 2:00 p.m.	<b>Clinical Proteomic Technology Assessment for Cancer</b> Daniel C. Liebler, Ph.D. Professor Vanderbilt University Medical Center
2:00 p.m. - 2:30 p.m.	<b>APT: the Analytical Proteomics Team</b> Fred E. Regnier, Ph.D. J.H. Law Distinguished Professor Purdue University
2:30 p.m. - 3:00 p.m.	<b>Measuring Cancer Biomarker Candidates by Targeted MS and Ab Enrichment</b> Steven A. Carr, Ph.D. Director, Proteomics Broad Institute of MIT and Harvard
3:00 p.m. - 3:30 p.m.	<b>CPTAC Team Leader Panel Discussion (Intra-Group Studies)</b>
3:30 p.m. - 3:45 p.m.	<b>Coffee Break and Poster Session</b>



4:15 p.m. - 5:15 p.m.

**Lessons Learned (Team Science): Clinical Proteomic Technologies for Cancer and International Leader Panel Discussion**

Moderator: Susan J. Fisher, Ph.D.

Steven A. Carr, Ph.D., *Director, Proteomics, Broad Institute of MIT and Harvard*

Paul Tempst, Ph.D., *Professor, Molecular Biology Program, Memorial Sloan-Kettering Cancer Center*

Daniel C. Liebler, Ph.D., *Professor, Vanderbilt University Medical Center*

Fred E. Regnier, Ph.D., *J.H. Law Distinguished Professor, Purdue University*

Henry Rodriguez, Ph.D., M.B.A., *Director, Clinical Proteomic Technologies for Cancer, National Cancer Institute, NIH*

Fredrik Pontén, M.D., Ph.D., *Director, Uppsala University Site for the Swedish Human Proteome Resource Project*

Myeong-Hee Yu, Ph.D., *Director, Korean Functional Proteomics Program*

Kenneth R. Evans, Ph.D., *President and CEO, Ontario Cancer Biomarker Network*

Steven I. Gutman, M.D., M.B.A., *Director, Office of In Vitro Diagnostic Device Evaluation and Patient Safety, Center for Devices and Radiological Health, FDA*

5:00 p.m. - 5:15 p.m.

**Summary and Action Items of Day 1**

Steven A. Carr, Ph.D.

*Director, Proteomics*

*Broad Institute of MIT and Harvard*

Henry Rodriguez, Ph.D., M.B.A.

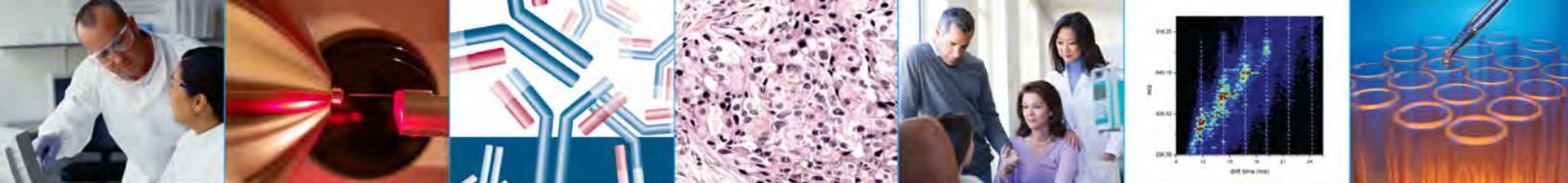
*Director, Clinical Proteomic Technologies for Cancer  
National Cancer Institute, NIH*

5:15 p.m.

**Meeting Adjournment**

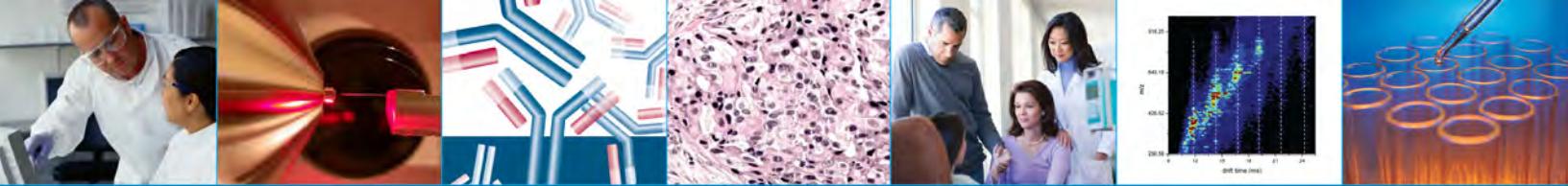
7:00 p.m. - 9:00 p.m.

**CPTAC Program Coordinating Committee Dinner**

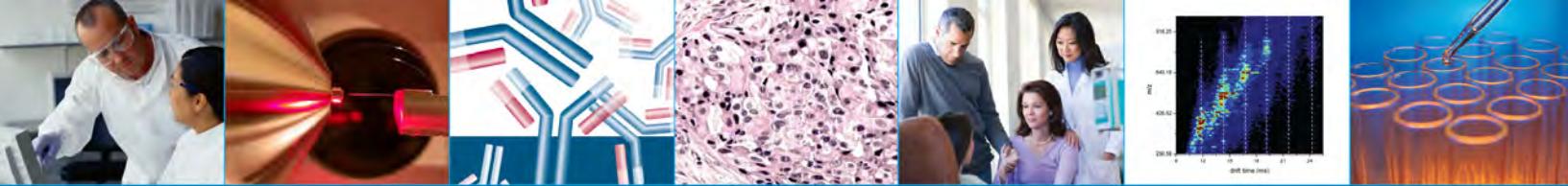


## **Day 2, October 25**

7:30 a.m. - 8:20 a.m.	<b>Registration</b>
8:20 a.m. - 8:25 a.m.	<b>Welcome – Review of the Agenda</b> Henry Rodriguez, Ph.D., M.B.A. Director, Clinical Proteomic Technologies for Cancer National Cancer Institute, NIH
8:25 a.m. - 8:35 a.m.	<b>Proteomic Reagents and Resource Core</b> Henry Rodriguez, Ph.D., M.B.A. Director, Clinical Proteomic Technologies for Cancer National Cancer Institute, NIH
8:35 a.m. - 8:45 a.m.	<b>Clinical Proteomic Technologies for Cancer Informatics Strategy</b> Chris Kinsinger, Ph.D. Program Manager, Clinical Proteomic Technologies for Cancer National Cancer Institute, NIH
8:45a.m. - 12:15 p.m.	<b>Advanced Proteomic Platforms Program and Computational Science Session</b> Moderator: Daniel C. Liebler, Ph.D. and Chris Kinsinger, Ph.D.
8:45 a.m. - 9:00 a.m.	<b>Proteomic Characterization of Alternate Splicing and cSNP Protein Isoforms</b> Nathan J. Edwards, Ph.D. University of Maryland, College Park
9:00 a.m. - 9:15 a.m.	<b>Enhancement of MS Signal Processing Toward Improved Cancer Biomarker Discovery</b> Dariya Malyarenko, Ph.D. College of William and Mary
9:15 a.m. - 9:30 a.m.	<b>Proteomic Phosphopeptide Chip Technology for Protein Profiling</b> Xiaolian Gao, Ph.D. University of Houston
9:30 a.m. - 9:45 a.m.	<b>Global Production of Disease-Specific Monoclonal Antibodies</b> Barry L. Karger, Ph.D. Northeastern University



9: 45 a.m. - 10:00 a.m.	<b>Analysis and Statistical Validation of Proteomic Datasets</b> Alexey I. Nesvizhskii, Ph.D. University of Michigan
10:00 a.m. - 10:15 a.m.	<b>Quantitative Methods for Spectral and Image Data in Proteomics Research</b> Timothy W. Randolph, Ph.D. Fred Hutchinson Cancer Research Center
10:15 a.m. - 11:00 a.m.	<b>Coffee Break and Poster Session</b>
11:00 a.m. - 11:15 a.m.	<b>Top-Down Mass Spectrometry of Salivary Fluids for Cancer Assessment</b> Joseph A. Loo, Ph.D. University of California Los Angeles
11:15 a.m. - 11:30 a.m.	<b>A New Platform to Screen Serum for Cancer Membrane Proteins</b> Daniel B. Martin, M.D. Institute for Systems Biology
11:30 a.m. - 11:45 a.m.	<b>Computational Tools for Cancer Proteomics</b> Katheryn A. Resing, Ph.D. University of Colorado at Boulder
11:45 a.m. - 12 noon	<b>New Proteomic Algorithms to Identify Mutant or Modified Proteins</b> David L. Tabb, Ph.D. Vanderbilt University
Noon - 12:15 p.m.	<b>A Proteomics Approach to Ubiquitination</b> Junmin Peng, Ph.D. Emory University
12:15 p.m. - 12:30 p.m.	<b>A Proteomics Platform for Quantitative, Ultra-High Throughput, and Ultra-Sensitive</b> Richard D. Smith, Ph.D. Battelle Pacific Northwest Laboratories
12:30 p.m. - 12:45 p.m.	<b>PICquant-An Integrated Platform for Biomarker Discovery</b> Dennis J. Templeton, Ph.D. University of Virginia



12:45 p.m. - 1:00 p.m.

**A Platform for Pattern-Based Proteomic Biomarker Discovery**

Denkanikota Mani, Ph.D.

Massachusetts Institute of Technology

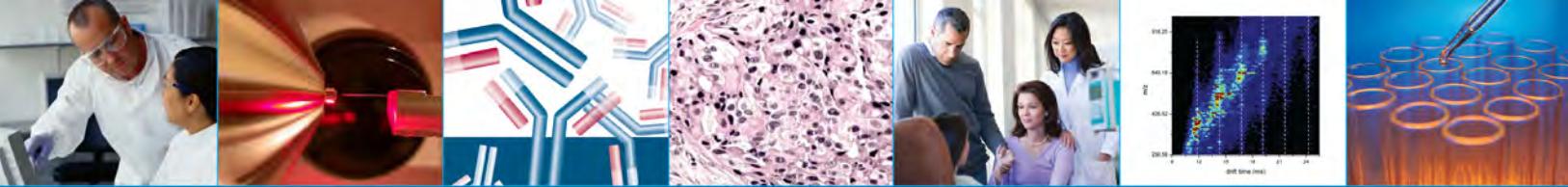
1:00 p.m. – 1:15 p.m.

**Closing Remarks, Action Items, and Adjournment**

Henry Rodriguez, Ph.D., M.B.A.

Director, Clinical Proteomic Technologies for Cancer

National Cancer Institute, NIH



## 4.8.2 2008 Annual Meeting Agenda



National Cancer Institute  
National Institutes of Health  
U.S. Department of Health and Human Services

### Clinical Proteomic Technologies for Cancer 2008 Annual Meeting

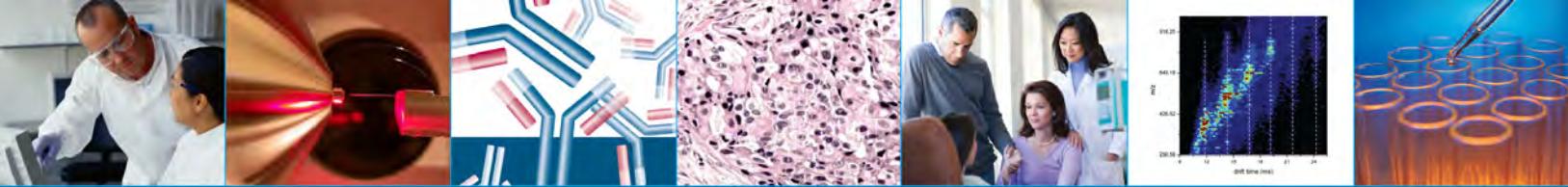
October 28-29, 2008

Hyatt Regency Cambridge  
Cambridge, MA

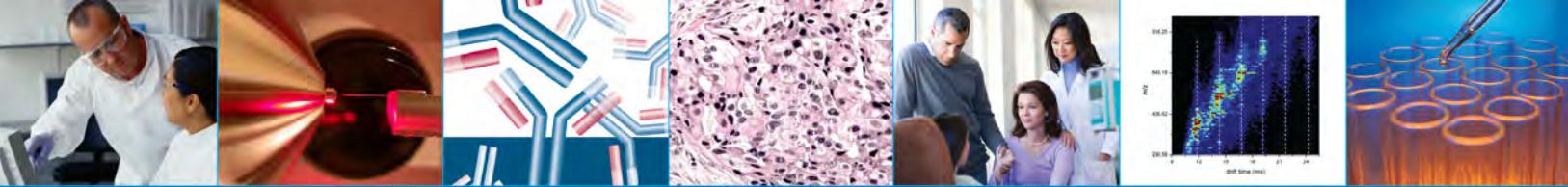
---

#### Day 1, Tuesday, October 28, 2008

- 7:30 a.m. – 8:00 a.m. **Registration and Morning Break**
- 8:00 a.m. – 8:10 a.m. **Welcome and Introduction**  
Henry Rodriguez, Ph.D., M.B.A.  
Director, Clinical Proteomics Technologies for Cancer  
National Cancer Institute, National Institutes of Health
- 8:10 a.m. – 8:50 a.m. **Keynote Lecture:  
Genomic Variation and Disease**  
David Altshuler, M.D., Ph.D.  
Professor of Genetics and Medicine  
Harvard Medical School/Massachusetts General Hospital  
Director  
Program in Medical and Population Genetics  
Broad Institute of MIT and Harvard
- 8:50 a.m. – 12:20 p.m. **Clinical Proteomics Technologies for Cancer (CPTC)  
and Innovative Molecular Analysis Technologies  
(IMAT) Program Joint Session: Fostering Collaborative  
Efforts**
- Co-Chairs:** Christopher Kinsinger, Ph.D.  
Clinical Proteomics Technologies for Cancer  
National Cancer Institute, National Institutes of Health



	J. Randy Knowlton Innovative Molecular Analysis Technologies Program National Cancer Institute, National Institutes of Health
8:50 a.m. – 9:15 a.m.	<b>CPTC Antibody Characterization Pipeline—Grand Opening</b> <i>Tara Hiltke, Ph.D.</i> <i>Clinical Proteomics Technologies for Cancer</i> <i>National Cancer Institute, National Institutes of Health</i>
9:15 a.m. – 9:40 a.m.	<b>Technology to Optimize SCFVs for Targeting Therapeutics</b> <i>Mark Federspiel, Ph.D.</i> <i>Mayo Clinic</i>
9:40 a.m. – 10:05 a.m.	<b>Incorporating IMAT-Developed Technology Into Another NCI Program</b> <i>Paul Tempst, Ph.D.</i> <i>Memorial Sloan-Kettering Cancer Center</i>
10:05 a.m. – 10:20 a.m.	<b>Break</b>
10:20 a.m. – 10:45 a.m.	<b>An Integrated Programmatic Environment to Support In-depth Proteomics and Phosphoproteomics Profiling and Studies of Gas Phase Chemistry of Peptides and Phosphopeptides</b> <i>Katheryn A. Resing, Ph.D.</i> <i>University of Colorado at Boulder</i>
10:45 a.m. – 11:10 a.m.	<b>Performance and Optimization of LC-MS/MS Platforms for Proteomic Analyses: An Interlaboratory Study</b> <i>Daniel Liebler, Ph.D.</i> <i>Vanderbilt University Medical Center</i>
11:10 a.m. – 11:35 a.m.	<b>Nanoscaled Detection of the Disease Proteome</b> <i>Nicholas Wang, Ph.D.</i> <i>Lawrence Berkeley National Laboratory</i>
11:35 a.m. – 12 noon	<b>Verification of Candidate Protein Biomarkers: Reproducibility of MRM-Based Assays</b> <i>Steven Hall, Ph.D.</i> <i>University of California, San Francisco</i>
12 noon – 1:30 p.m.	<b>Lunch and Poster Session</b>



1:30 p.m. – 3:10 p.m.

**Session 1: CPTC Technologies**

**Chair:** Susan Fisher, Ph.D.

*University of California, San Francisco*

1:30 p.m. – 1:55 p.m.

**Multiarray Assay Platform for Cancer Biomarker Validation and Discovery**

*John H. Kenten, Ph.D.*

*Meso Scale Diagnostics, LLC*

1:55 p.m. – 2:20 p.m.

**Improving the Sensitivity of Peptide Identification With Metasearch and Machine Learning**

*Nathan Edwards, Ph.D.*

*Georgetown University*

2:20 p.m. – 2:45 p.m.

**Automated Multianalyte Profiles (MAPs) for Early Detection of Cancer-Related Proteins**

*Ralph McDade, Ph.D.*

*Rules-Based Medicine, Inc.*

2:45 p.m. – 3:10 p.m.

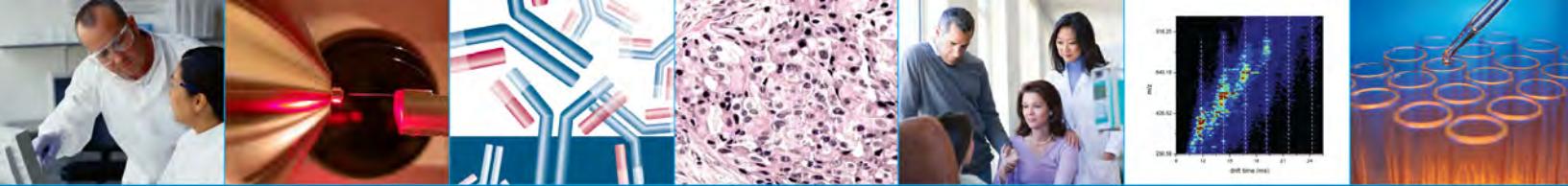
**Significance Analysis of Spectral Count Data in Label-Free Shotgun Proteomics**

*Alexey I. Nesvizhskii, Ph.D.*

*University of Michigan*

3:10 p.m. – 3:40 p.m.

**Break and Poster Session**



3:40 p.m. – 5:20 p.m.

**Session 2: CPTC Technologies**

**Chair:** *Daniel Liebler, Ph.D.*

*Vanderbilt University Medical Center*

3:40 p.m. – 4:05 p.m.

**Discovery and Validation of Lung Cancer-Specific Monoclonal Antibodies**

*Lazlo Takacs M.D., Ph.D., Sc.D.*

*Biosystems International*

4:05 p.m. – 4:30 p.m.

**Combining PCR Amplification Technology and Mass Spectrometry for Sensitive Detection of Proteins**

*Charles Cantor*

*Sequenom, Inc.*

4:30 p.m. – 4:55 p.m.

**Computational Reassembly of Fractionated Samples for Biomarker Discovery using PEPPeR**

*D.R. Mani, Ph.D.*

*Broad Institute of MIT and Harvard*

4:55 p.m. – 5:20 p.m.

**High-Throughput, High-Sensitivity Proteomics Platform for Improved Biomarker Discovery and Verification**

*Richard D. Smith, Ph.D.*

*Pacific Northwest National Laboratory*

5:20 p.m. – 5:30 p.m.

**Wrap-Up and Adjournment**

*Henry Rodriguez Ph.D., M.B.A.*

*Director, CPTC*

*NCI, NIH*

5:30 p.m. – 8:30 p.m.

**CPTC Boston Fair and Poster Session**

**Day 2, Wednesday, October 29, 2008**

7:30 a.m. – 8:00 a.m.

**Registration and Morning Break**

8:00 a.m. – 8:10 a.m.

**Welcome**

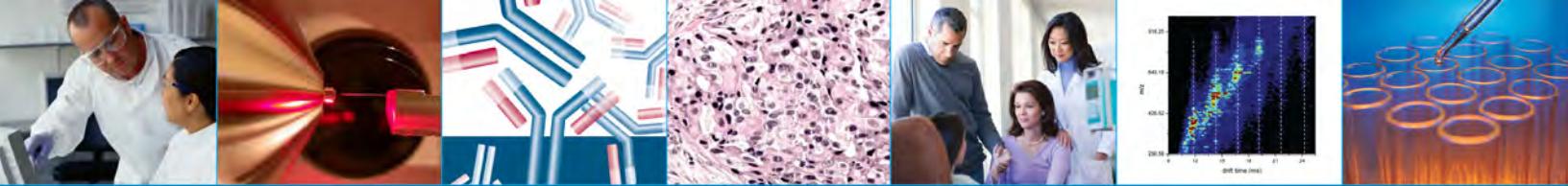
*Steven Carr, Ph.D.*

*Broad Institute of MIT and Harvard*

8:10 a.m. – 8:50 a.m.

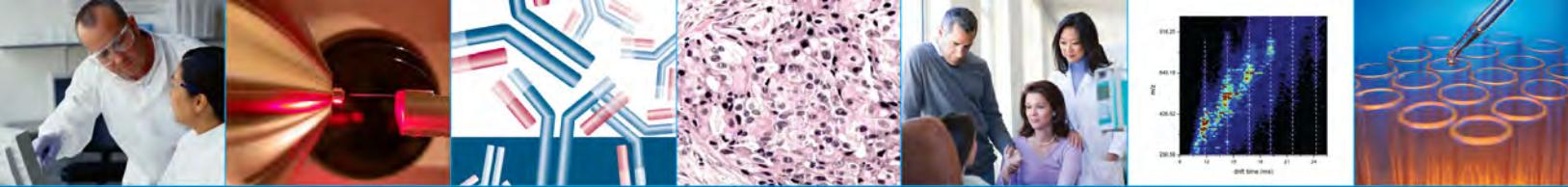
**Convergence of Genomic, Proteomic, and Metabolomic Technologies To Unravel the Function of Mitochondria in Human Disease: Lessons Learned and What Is Still Needed**

*Vamsi Mootha, M.D.*



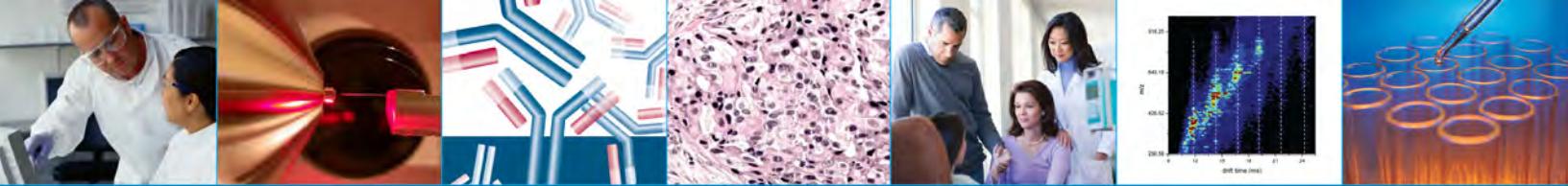
*Harvard Medical School*

- 8:50 a.m. – 10:05 a.m. **Session 1: CPTC Next-Generation Computational Tools and Platforms**  
**Chair:** *Fred E. Regnier, Ph.D.*  
*Purdue University*
- 8:50 a.m. – 9:15 a.m. **Characterization of Polyubiquitin Chain Structure by Middle-Down Mass Spectrometry**  
*Junmin Peng, Ph.D.*  
*Emory University School of Medicine*
- 9:15 a.m. – 9:40 a.m. **DirecTag Infers Partial Peptide Sequences From MS/MS Scans**  
*David L. Tabb, Ph.D.*  
*Vanderbilt University School of Medicine*
- 9:40 a.m. – 10:05 a.m. **An Immunochromatographic Array Platform for Cancer Proteomics**  
*Fred E. Regnier, Ph.D.*  
*Quadraspec, Inc.*
- 10:05 a.m. – 10:20 a.m. **Break**
- 10:20 a.m. – 12 noon **Session II: CPTC Next-Generation Computational Tools and Platforms**  
**Chair:** *Paul Tempst, Ph.D.*  
*Memorial Sloan-Kettering Cancer Center*
- 10:20 a.m. – 10:45 a.m. **Comparative Proteomics for the Discovery of Ubiquitin-Ligase Proteins**  
*Timothy Randolph, Ph.D.*  
*University of Washington*
- 10:45 a.m. – 11:10 a.m. **Isolation and Characterization of SCFV Antibodies Against Cancer Antigens Using Novel Display Library**  
*Jiwu Wang*  
*Allele Biotechnology and Pharmaceuticals*
- 11:10 a.m. – 11:35 a.m. **Enhancement of Mass Spectrometry Signal Processing Toward Improved Cancer Biomarker Discovery**  
*Dariya Malyarenko, Ph.D., M.S.*  
*College of William and Mary*



11:35 a.m. – 12 noon

**MAZIE: A Mass and Charge Inference Engine To Enhance Database Searching of Tandem MS Spectra**  
*Dennis J. Templeton, M.D., Ph.D.  
University of Virginia Health System*



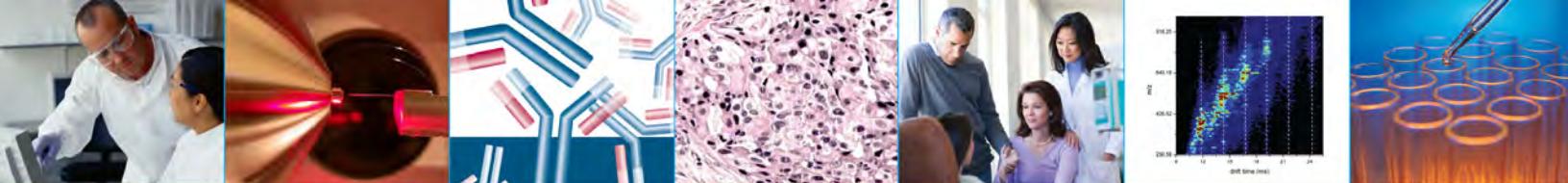
## 4.9 NCI-FDA Memorandum of Understanding

### MEMORANDUM OF UNDERSTANDING

NATIONAL CANCER INSTITUTE

AND THE

U.S. FOOD AND DRUG ADMINISTRATION



## National Cancer Institute – Food and Drug Administration

### Memorandum of Understanding

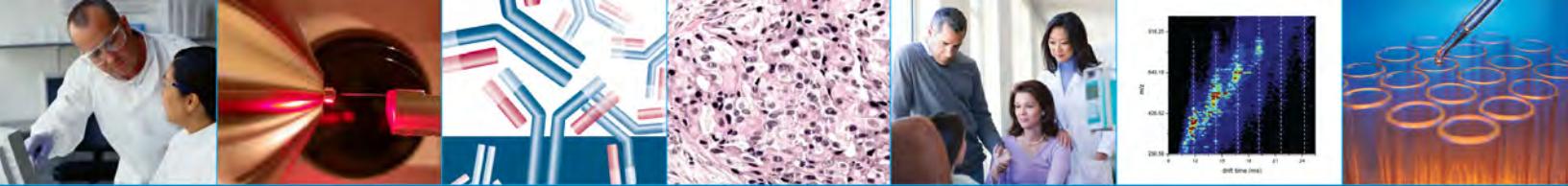
The National Cancer Institute and the Food and Drug Administration, within the Department of Health and Human Services, agree that it is in the best interests of both agencies to develop a partnership that leverages each agency's core expertise and resources. These efforts will serve to facilitate new science and technology initiatives to advance U.S. strengths and international programs, as appropriate, in health and medical science and technology fields. Additional benefits include improved safety and effectiveness via targeted therapies based on patient proteomic profiles, decreased death and suffering due to cancer and improved public health. The NCI and FDA have mutual interests in understanding the molecular biology of cancer as it relates to diagnostic measurements, product development, and therapeutic treatment in the course of their respective individual missions. It is within the scope of these interests, as they pertain to proteomics technologies and proteomics applications in clinical cancer diagnostics and cancer therapeutics development, that this memorandum is developed.

#### 1. PURPOSE AND SCOPE

The purpose of this Memorandum of Understanding is to extend formal scientific and programmatic collaborations between the NCI and FDA. Such cooperation will facilitate the two agencies fulfilling their unique missions. The scope of the collaborations covered by this agreement includes proteomics science and technology applications such as instrument/technology validation, informatics, biological sample preparation, diagnostics, and discovery and validation of biomarkers for cancer development, treatment, and response,. This MOU sets forth the basic principles and guidelines under which the agencies intend to work together to foster partnership in: research and education; the exchange of ideas, information, and data; the development and use of proteomics technologies in clinical applications; mutual collaborations with industrial, non-profit, academic, and other government institutions and agencies; and other efforts to further the advancement of and application of proteomics technologies to improve patient response to treatment, reduce the pain and suffering of cancer and improve the standard of living and quality of life.

Achievements in science and technology now require a higher level of integration, particularly in the development of interdisciplinary research teams. The programmatic strengths of the NCI and FDA in the areas of molecular biology, technology development, protein chemistry, analytical validation and clinical trials designs offer the opportunity for synergy that will accelerate proteomics technology development and application in clinical settings.

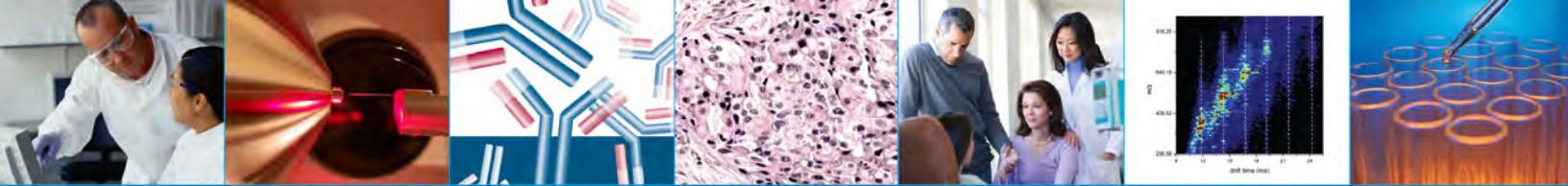
Although with a different emphasis, each institution has existing programs and resources for facilitating technology development. The NCI has recently launched the Clinical Proteomics Technology Initiative for Cancer to specifically address limitations and challenges in applying clinical proteomics to alleviate the cancer burden. This program



will advance the field of proteomics to allow data and results to be compared and compiled to better understand changes in the proteome as they are reflected in cancer and will serve as the main conduit for exchange of information between these two agencies pursuant to this MOU.

The FDA's Office of In Vitro Diagnostics in the Center for Devices and Radiological Health (CDRH) was created to handle all aspects of in vitro diagnostic approval, postmarket problems, and compliance to ensure that approved devices are based on valid scientific evidence of safety and that they also provide evidence for effectiveness and clinical utility. For the integration of new technologies and molecular-based diagnostics, the development of appropriate measurement methods, standard protocols, and analysis platforms is essential. The FDA's Center for Biological Evaluation and Research (CBER) was created to ensure the safety, purity, potency, and effectiveness of biological products including vaccines, blood and blood products, and cells, tissues, and gene therapies for the prevention, diagnosis, and treatment of human diseases, conditions, or injury.. Proteomics technologies are intrinsically related to the development and monitoring of therapies for safety, potency and efficacy, and will facilitate the availability of critical medical therapies to save lives and improve human health. CBER monitors numerous attributes of a product for its life time. The FDA's Center for Proteomics and Center for Toxicoinformatics within the National Center for Toxicological Research (NCTR) conduct research in support of the FDA's current and future regulatory needs. The Center for Proteomics currently has a research emphasis in the development of mass spectrometry based proteomic methods as well as in the development of proteomic informatics tools to aid in data analysis (in collaboration with the Center for Toxicoinformatics). The Center for Toxicoinformatics continues development of technologies to incorporate, manage, and analyze data from proteomic experiments to improve sponsor data submission channels. This MOU provides a framework that will enable and encourage the sharing of knowledge and support the formation of research teams to solve complex problems in the area of each agency's mission.

The parties agree to collaborate through working groups and partnerships to develop strategic plans, set priorities, and leverage resources and expertise from multiple sources, including the private sector. Both organizations will adhere to all appropriate protocols and procedures to ensure there are no conflict of interest issues existing in these initiatives, toward the goal of facilitating the development of proteomics technologies in clinical research and applications that constitute novel tools, more effective diagnostics, better characterization of novel products and molecularly based personalized cancer therapeutics. This will be done, in part, by establishing a framework for effective assessment and evaluation of existing and emerging proteomics technologies. This MOU sets forth the framework for collaboration among the Parties and for pursuing specific collaborative projects that may involve additional partners and will be implemented through separate agreements. The Parties anticipate that concepts developed and activities undertaken under the auspices of this MOU may lead to partnerships that will be implemented through separate agreements.



## 2. AUTHORITY

This MOU is authorized under Section 301 of the Public Health Service Act which authorizes NIH to cooperate with public authorities and scientific institutions.

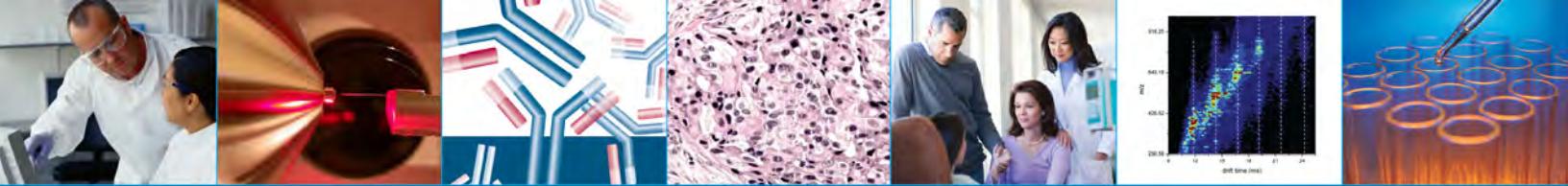
## 3. BACKGROUND

The NCI is the Federal government's principal agency for the support of cancer research and training. In this context, the NCI coordinates the Clinical Proteomic Technology Initiative for Cancer, which supports research, training, data collections and analysis, technology assessment, and other programs with respect to understanding and analyzing the field of proteomics in order to improve technological capabilities in early detection, diagnostics, prevention, therapeutic monitoring, and treatment of cancer. The NCI has identified the development of high impact technologies to support cancer research as a priority area for enhanced commitment.

FDA's role is to regulate medical products and to develop new policies to allow commercial marketing and distribution of new, safe and effective technologies and products. Consequently, the FDA has launched the Critical Path Research Initiative to identify, develop, and apply state-of-the-art genomics and proteomics technologies to product characterization and clinical analytical design for faster, more predictable development and regulatory approval of safe and effective innovative medical products for cancer to enhance Public Health. Cutting edge proteomics technologies are being planned for the development of new cancer diagnostic and therapies, identification of biomarkers for product characterization and to promote application of proteomics in validation of preclinical and clinical data for FDA regulated products. Those products will be used as new cancer diagnostics to prevent and monitor disease and promote public health.

The complementary and synergistic roles of the NCI and the FDA facilitate the increased cooperation planned by this MOU. This MOU will facilitate the NCI improving its transfer of science, technology, and engineering discoveries from its funded institutions under the Clinical Proteomics Technology Initiative for Cancer to the marketplace and clinic through linkage with the analytical, product characterization and clinical validation expertise at the FDA. Such a close collaboration will help promote the accomplishment of the goals of both agencies as summarized in their respective strategic platforms.

An important area of mutual interest is the application of proteomics technologies to disease diagnostics, product characterization, safety assessment, and therapeutic monitoring. Both agencies depend on the reproducibility of results to ensure accurate clinical applications. In order for the field of proteomics to move forward, both agencies would benefit from a mutual agreement of analytical, product characteristics, and clinical validation of proteomic technologies, proteomic data submission requirements, and the subsequent results and interpretations of such data.



Collaborative efforts will be taken by both the NCI and the FDA to incorporate studies that will support analytical, product and clinical validation concerns for evaluating data to support claims for in vitro diagnostic products, biomarkers, and therapies that will impact regulatory relevance of drugs, biologics, and devices based on proteomics data; develop joint educational seminars and workshops to make the scientific community aware of the research in applied and clinical proteomics; educate the NCI and its funded institutions on the current requirements, parameters, and concerns in regulatory data submission; and work in line with industrial and academic partners to reduce redundant studies and improve the efficiency of technology validation for pre-clinical, product development, and clinical applications.

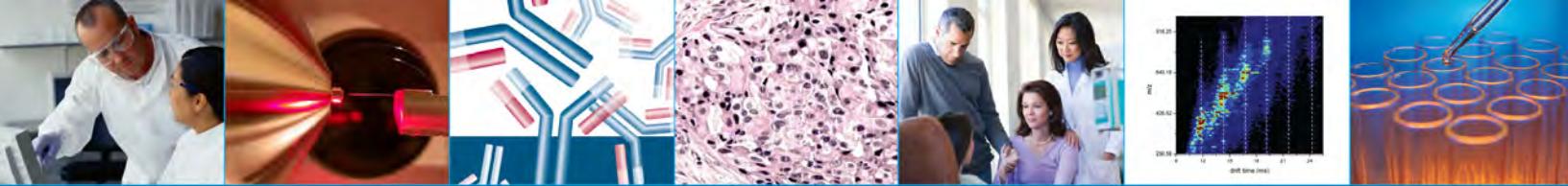
#### 4. COLLABORATION

This MOU records the intent of the parties to collaborate in areas including but not limited to the following aspects of proteomics: sample collection, preparation, storage, and processing, bioinformatics and data analysis, diagnostic assay development, and discovery and validation of biomarkers and surrogate biomarkers of cancer development and drug response, including standardization among technology platforms and assay standards development.

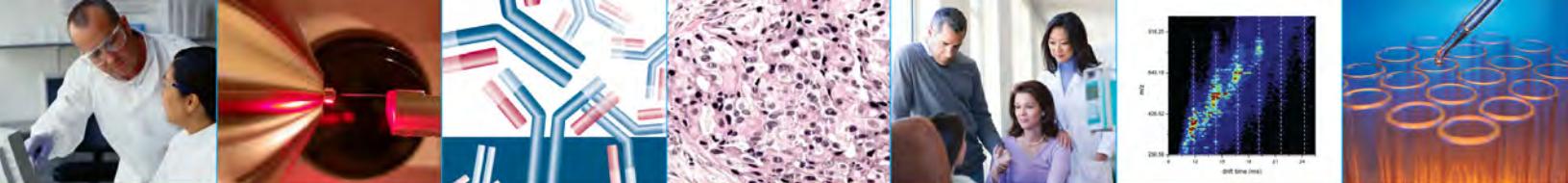
Any collaboration between NCI and FDA is expected to capitalize on the core strengths of each agency. It is expected that the NCI will have the lead role in cancer research activities and the FDA will have the lead role on issues regarding analytical validation of diagnostic and therapeutic products, product characteristics validation, and data submission for in vitro diagnostics and therapeutics evaluations.

This MOU is intended to provide an enabling mechanism for coordination and cooperation whenever appropriate and mutually beneficial.

- a. Cooperation may include but is not limited to the specific areas identified below:
  - i. Development of guidelines for necessary requirements for data submission and analysis for proteomics data repositories that is mutually beneficial to the NCI and the FDA. These requirements will be intended to meet the needs of the NCI for the purposes of technology assessment, validation, and experimental reproducibility and the needs of the FDA for the purposes of diagnostic and drug submissions.
  - ii. Development of joint educational material and programs in which the NCI and FDA detail the most current standards and practices for proteomics technologies and applications
  - iii. Formation of NCI, FDA, academic, industry and other government agency partnerships to improve proteomics standard operating procedures
  - iv. Mutual use of facilities, software, algorithms, and data repositories and collaboration on development of new facilities as appropriate
  - v. Cooperation to facilitate and enhance extramural research and development activities by either agency
  - vi. Cooperation through the exchange of agency personnel, expertise, scientific and technical information, data, and publications



- vii. Joint publicity of mutually reinforcing activities, publications, and research results, including hyperlinks to each others' programs on their websites
- viii. Mutual assistance in program planning, and in the review of research development projects and proposals
- ix. Inclusion of representatives from each agency in workshops, working groups, seminars, and other related activities.
- b. To pursue the collaboration described above the parties agree to use the following framework for implementing collaborations:
- i. Both the NCI and FDA will identify coordinators to implement and manage this MOU. The Director of the NCI's Clinical Proteomic Technologies Initiative for Cancer and the Director of the FDA's Office of In Vitro Diagnostics in the Center for Devices and Radiological Health or the designee, will be the primary representative to coordinate the activities from their respective agencies for the purposes of this MOU. The coordinators shall meet on a regular basis to discuss activities conducted under the MOU, review all aspects of implantation, and plan future directions of programmatic interaction and cooperation, and report to signatories annually.
  - ii. Concepts or ideas for developing collaborations or activities involving joint projects or integrated approaches to conducting science or technology development will be formally presented by submission of concepts to the coordinators of the MOU.
  - iii. Representatives from each agency will meet quarterly to review progress and address new opportunities for collaboration. Technical and programmatic advisory working groups made up of NCI and FDA Centers' employees may be assembled to make formal recommendations for collaboration.
  - iv. Industry and scientific organizations' representatives for relevant proteomic technologies may interact with the partners through guidelines established and accepted by each agency's representative. The guidelines will be established in keeping with the policies for both organizations.
  - v. The coordinators shall seek to resolve any dispute concerning the MOU through good faith discussions with the coordinators and ombudsman from the respective organizations when needed.
- c. Cost and funding
- i. Costs associated with the participation of NCI and FDA shall be borne subject to the availability of appropriated funds and designated personnel by each agency, or the approval of other sources of funding. Funding and resources for each significant activity undertaken pursuant to this MOU shall be arranged in accordance with the applicable mechanisms. The MOU will provide a framework for the transfer of resources or personnel from one Agency to the other if needed using appropriate mechanisms (e.g., Interagency Agreements).
  - ii. The NCI and FDA agree that this MOU does not commit either agency to make specific levels of financial or personnel support or to provide specific laboratory or office space for programs relevant to this MOU. The provision of such support will be based upon available resources and provided in accordance with the



rules, regulations, and laws under which FDA operates and the policies of the NCI and NIH.

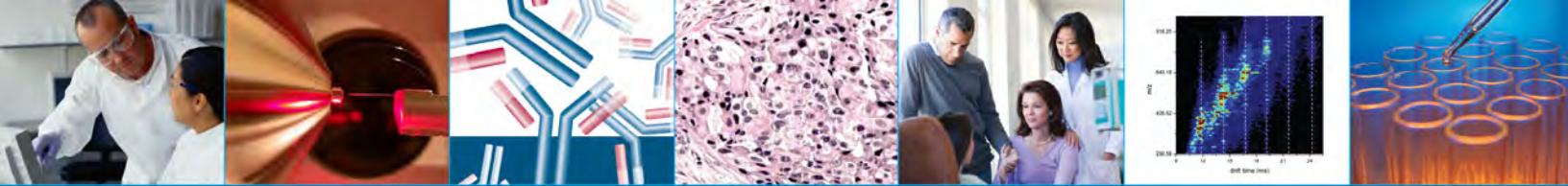
## 5. GENERAL PROVISIONS

Proprietary and/or nonpublic information will not be disclosed under this MOU, unless such disclosure is governed by the appropriate confidentiality disclosure agreements, to the extent such disclosure is permitted by law.

Materials being analyzed/studied under the terms and conditions of this MOU may be shared among the Parties and such transfers will be under separate Material Transfer Agreements (MTA) or other appropriate agreements as permitted by law.

Rights to intellectual property developed during the course of research under this agreement will be addressed in separate project-specific implementing agreements.

Any notice or other communication required or permitted under this MOU will be in writing and will be deemed given as of the date it is received and accepted by the receiving party.



## 6. CONTACTS

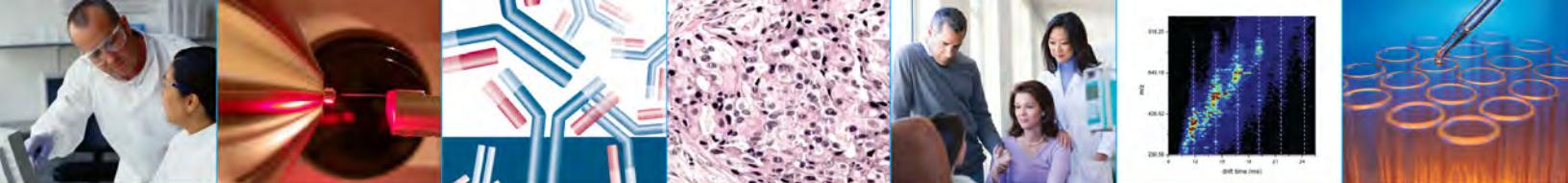
Notices or formal communications pursuant to this MOU should be sent to:

For FDA: Francis Kalush, Ph.D.  
OIVD, CDRH  
2098 Gaither Rd  
HFZ-440  
Rockville, MD. 20850  
Telephone: (240) 276-0996

For NCI: Henry Rodriguez., Ph.D.  
Office of the Director, NCI  
31 Center Drive  
MSC 2580 - Room 10A52  
Bethesda, MD 20892  
Telephone: (301) 496-1550

## 7. TERM, MODIFICATIONS AND TERMINATION

- a. This MOU constitutes the entire agreement among the Parties pertaining to proteomics at the NCI's Center for Strategic Science and Technology Initiatives and FDA.
- b. There are no representations, warranties, agreements or understandings, express or implied, written or oral between the Parties hereto relating to the subject matter of this MOU that are not fully expressed herein.
- c. No supplements, amendments or modifications to this MOU shall be binding unless executed in writing, with thirty (30) days advance notice, and by mutual consent of the Parties; such modifications are to take the form of amendments.
- d. This MOU, when accepted by the Parties, will have an effective date from the date of the last to sign and will remain in effect for five (5) calendar years from the effective date, unless modified or terminated. Either party may terminate this MOU at any time provided a 90-day written notice is provided to the other agency and appropriate steps are taken to ensure an orderly termination of joint activities.



AGREED TO:

UNITED STATES FOOD AND DRUG ADMINISTRATION

BY: \_\_\_\_\_  
Signature of authorized representative \_\_\_\_\_ Date

Janet Woodcock, M.D. \_\_\_\_\_  
Print Name \_\_\_\_\_ Chief Medical Officer  
\_\_\_\_\_  
U.S. Food and Drug Administration (FDA)  
Title

BY: \_\_\_\_\_  
Signature of authorized representative \_\_\_\_\_ Date

Andrew C. von Eschenbach, M.D. \_\_\_\_\_ Commissioner,  
Print Name \_\_\_\_\_ U.S. Food and Drug Administration FDA  
Title

NATIONAL CANCER INSTITUTE

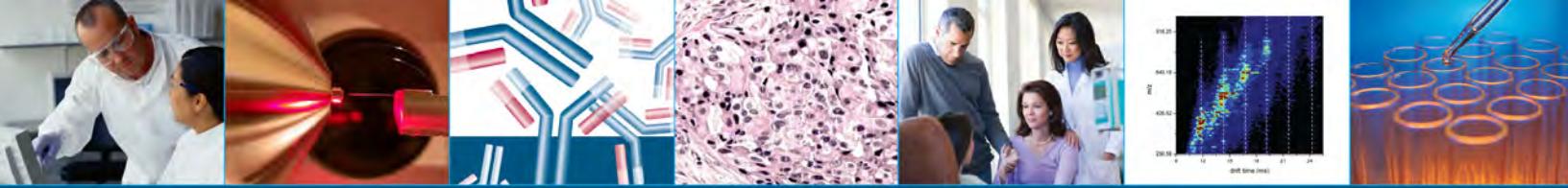
BY: \_\_\_\_\_  
Signature of authorized representative \_\_\_\_\_ Date

Anna Barker, Ph.D. \_\_\_\_\_  
Print Name \_\_\_\_\_ Title

BY: \_\_\_\_\_  
Signature of authorized representative \_\_\_\_\_ Date  
John E. Niederhuber, M.D  
Print Name Director, National Cancer Institute  
National Institutes of Health  
Title

EFFECTIVE DATE

This MOU is effective on the date of the last to sign.



## 4.10 NCI-FDA Interagency Oncology Task Force Molecular Diagnostics Workshop Agenda

### Agenda

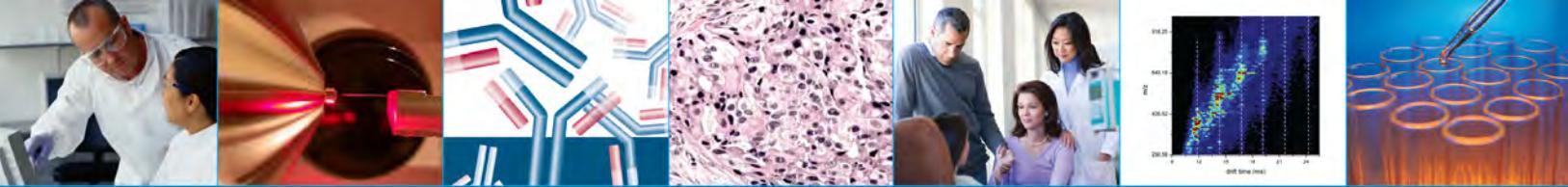
#### NCI-FDA Interagency Oncology Task Force Molecular Diagnostics Workshop

Hyatt Regency in Cambridge, MA

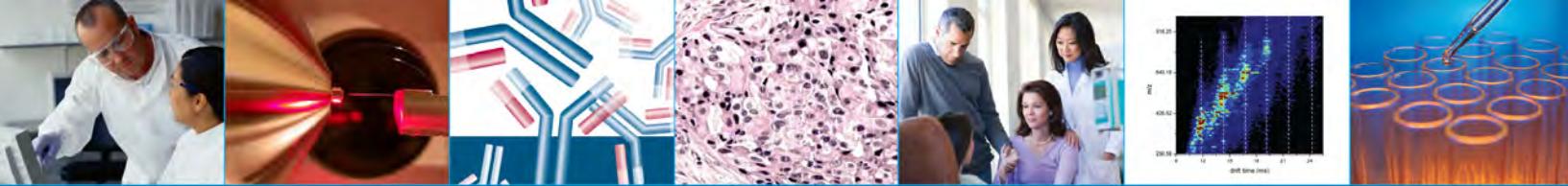
Oct. 30, 2008

---

7:45 a.m. – 8:00 a.m.	<b>Welcome and Introduction</b> <i>Henry Rodriguez (NCI) / Elizabeth Mansfield (FDA)</i>
8:00 a.m. – 8:45 a.m.	<b>FDA: Overview and Recommendations for IVDs</b> <i>Elizabeth Mansfield (FDA)</i>
8:45 a.m. – 10.00 a.m.	Case Study (Moderator: D. Ransohoff, UNC) <b>Approved Biomarkers Based on Microarray Platform (MammaPrint)</b> <i>Reena Philip (FDA)</i>
10:00 a.m. – 10:30 a.m.	Case Study (Moderator: L. Anderson, PPI) <b>Newborn Metabolite Screening Test Systems Using Tandem Mass Spectrometry</b> <i>Elizabeth Mansfield (FDA)</i>
10:30 a.m. – 10:45 a.m.	<b>Break</b>
10:45 a.m. – 11:45 a.m.	Case Study (Moderator: S. Carr, Broad) <b>Colon Cancer Biomarker Pipeline</b> <i>Dan Liebler (Vanderbilt)</i>
11:45 a.m. – 1:00 p.m.	Case Study (Moderator: S. Carr, Broad) <b>Mass Spectrometry-Based Verification of Candidate Plasma Protein Biomarkers for Early Detection of Severe Preeclampsia</b> <i>Mike McMaster (UCSF)</i>
1:00 p.m. – 1:45 p.m.	<b>Working lunch</b>
1:45 p.m. – 2:45 p.m.	Case Study (Moderator: S. Fisher, UCSF) <b>Enzymatic Activities as Biomarkers for Cancer: A Case Study on Blood-based Exopeptidases</b> <i>Paul Tempst (MSKCC)</i>



- 2:45 p.m. – 3:00 p.m. Break
- 3:00 p.m. – 4:15 p.m. Case Study (Moderator: S. Skates, MGH)  
**Immunological Array and Liquid Chromatography . Biomarker Assay Systems**  
*Fred Regnier (Purdue)*
- 4:15 p.m. – 5:00 p.m. **Succinct and Itemized Discussion of Lessons Learned**  
*Moderator: Larry Kessler (FDA)*
- 5:00 p.m. – 5:30 p.m. **Action Items**  
*Henry Rodriguez (NCI) / Elizabeth Mansfield (FDA)*



#### 4.11 NHLBI Adoption of CPTAC Verification Technology

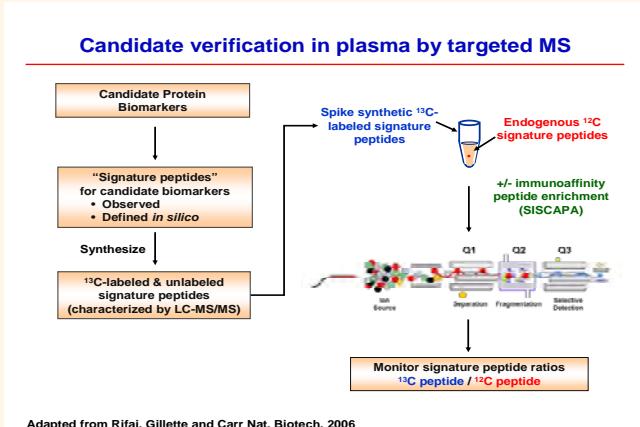
In their search for valid biomarkers for acute coronary syndromes, scientists from the National Heart, Lung and Blood Institute (NHLBI) are adopting verification technology introduced by CPTAC scientists that bridges biomarker discovery and clinical validation.

## Examples of Work-in-Progress: Development of Novel Analytical Tools for Early Biomarker Characterization:

Once a list of potentially informative cardiovascular biomarkers is compiled, we begin to characterize the candidates for which there are minimal or no data in acute coronary syndromes in humans. Before embarking on large-scale validation studies, basic knowledge is required regarding kinetics and relation to different coronary artery disease states. Our group is actively engaged in investigation using targeted mass spectrometry (MS) as a bridge from discovery to validation, in studies led by Terri Addona and Steve Carr at the Broad Institute in collaboration with the Gerszten and Sabatine groups. This aspect of our work also synergizes with our development of a targeted MS metabolomics platform (Sabatine, Gerszten, et. al. Metabolomic Identification of Novel Biomarkers of Myocardial Ischemia, *Circulation* 2005; 112: 3868-75, Lewis, Carr, Sabatine, Gerszten et. al, *JCI* 2008;118:3505-12). Our working hypothesis is that targeted MS will accelerate the validation process when ELISA reagents are not available. Furthermore, we are working on protocols to use targeted MS for the simultaneous assessment of potentially dozens of analytes in a given sample in a high throughput, low-cost manner. An overview of our targeted LC-MS pipeline is shown in the figure. In our pipeline are analytes for which ELISA are not readily available (eg., IL-33, MRP-14). Additional proteins from our ongoing proteomics discovery work are also in our targeted MS pipeline. (SISCAPA: Stable Isotope Standards and Capture by Anti-Peptide Antibodies; MRM: Multiple Reaction Monitoring)

For initial validation, we are using two particularly informative cohorts for characterization of markers: patients experiencing planned myocardial ischemia during exercise stress testing and patients experiencing planned myocardial infarction (PMI) during alcohol septal ablation for hypertrophic cardiomyopathy. In both, the con-

trolled, timed nature of the myocardial insult permits samples to be obtained both before and after the injury, allowing each patient to serve as his or her own biological control. Additional important cohorts include small phase II studies of patients with acute coronary syndromes in which a wealth of adjunctive hemostatic, angiographic, hemodynamic, and electrocardiographic data are available; cohorts of patients with stable coronary artery disease; and healthy controls.



Adapted from Rifai, Gillette and Carr Nat. Biotech. 2006

## Recent Awards to Faculty and Staff

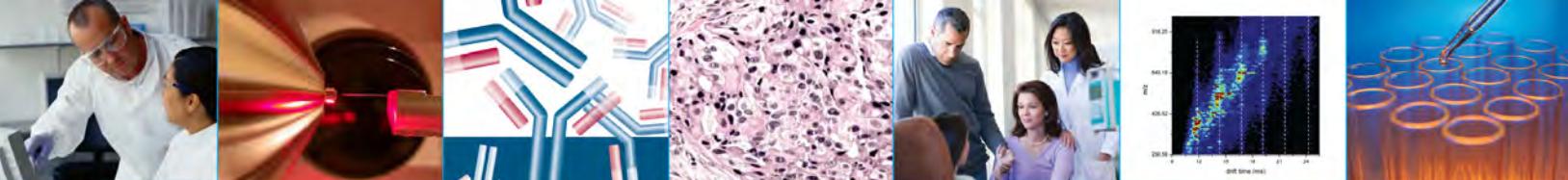
### Dr. Jessica L. Mega

Samuel A. Levine Young Clinical Investigator Award, American Heart Association  
National Clinical Research Grant, American Heart Association

### Dr. Michelle O'Donoghue

Women in Cardiology Trainee Award for Excellence, American Heart Association  
American College of Cardiology Scientific Session Bristol-Myers Squibb Travel Award  
Finalist, Best Poster Awards Competition, American College of Cardiology Scientific Session





## 4.12 Trade Media Article Reprints

### **Sharing the Wealth of Data**

*Scientific American worldVIEW*  
June 2009

### **Proteomics: a long and winding road to medical diagnostics**

*Expert Opinion on Medical Diagnostics*  
May 2009

### **Commentary: Complementary fields of proteomics and genomics are cornerstone of personalized medicine**

*Drug Discovery News*  
March 2009

### **Clinical Proteomics Heads Into Real World**

*Genetic Engineering and Biotechnology News*  
March 2009

## Society & Culture

# Sharing The Wealth Of Data

Combining knowledge is fundamental to innovation. Doing it right requires new technologies, policies and ways of interacting  
by Mike May

May 2009



(Illustrations by Brian Stauffer)

By the late 1990s, avalanches of data were pouring into GenBank—the genetic sequence database operated by the U.S. National Institutes of Health. In 1999 GenBank contained about two billion base pairs. That number

jumped to 11 billion in 2000, nearly 45 billion in 2004 and nearly 86 billion by 2008. As that database and many others grow, it becomes increasingly valuable to share knowledge, but doing so becomes ever more difficult. Speed makes up part of the problem, but policy hurdles also block the way.

Around 2003 the speed of sharing data concerned Wu Feng, then at Los Alamos National Laboratory and now an associate professor in the departments of computer science and electrical & computer engineering at Virginia Tech in Blacksburg. Feng knew that the number of bases in GenBank was growing faster than the ability to search them, especially with the popular basic local alignment search tool, better known simply as BLAST. So Feng and his colleagues created mpiBLAST, which lets multiple computer processors tackle the same sequencing query as a team. This new software made sequence searches faster—often several orders of magnitude faster—but Feng and his

colleagues would soon be searching for more ways to increase the speed of data sharing.

Meanwhile many other researchers pursued different approaches to connecting biological and biomedical information around the world. Although making these data connections demands solving technological and sociological challenges, the results can change the approach to basic research and even the business of biotechnology.

### Sharing Spreads More than Data

"Today more than ever, researchers recognize the impact of sharing," says Henry Rodriguez, director of clinical proteomic technologies for cancer (CPTC) at the U.S. National Cancer Institute (NCI). He adds, "Advances in science and healthcare are made possible through widespread and barrier-free access to research and the data produced by that research."

In fact Rodriguez sees at least five ways that data sharing benefits research. First, data sharing encourages open scientific enquiry. "This lets conclusions from research be validated or refuted by peers, and that adds more strength to the results," Rodriguez explains. Second, sharing data from past experiments triggers new ones. As Rodriguez says, "Existing data can lead to new insights that the first investigator might not have recognized." He adds, "Programs in genomics

and all of the ‘omics are producing vast amounts of data, but connecting the data and extracting knowledge from the data are critical." Third, Rodriguez points out that making data openly available creates huge test sets that can be used to assess the quality of new informatics software. Fourth, combining information creates data sets that cannot be generated by any individual. "Putting it all together is the key," Rodriguez says. Last, he believes that sharing data openly reduces unnecessary duplication. "Some duplication provides rigor," Rodriguez says, "but accessing data from others can also push science further."

In part the very nature of biotechnology demands data sharing. "Almost by definition," says Kenneth H. Buetow, associate director for bioinformatics and information technology at NCI, "biotechnology and biomedicine are international enterprises." He adds, "There are immediate challenges from that globalization, especially how to get continuity of information and connecting information."

Nonetheless, some information in biotechnology, such as proprietary data generated inside biotechnology and pharmaceutical companies, will never be readily released—at least not right after it's collected. But that is not necessarily the bulk of biotech information. "There are tons of resources nowadays," says Buetow, "that are precompetitive. So while this information is not necessarily proprietary, it is a disadvantage if a company cannot access the information and has to generate it." For example he points out that genome-wide association studies could be helpful to many researchers—in basic research and business—even though the commercial value is largely limited. "That information should be shared on a broad scale," Buetow says. "It is invaluable."

Still, Buetow knows that some information will not be made open to everyone. For example a pharmaceutical company is not going to openly share information about binding between a candidate drug and a

disease target. But even that information will need to be shared inside the company. Likewise in a multinational clinical trial, data might be shared between a pharmaceutical company and a contract research organization or local physicians collecting data. As Buetow says, "Some data need to be bound for intellectual property reasons or through licensing, but I would argue that even that needs to be shared. The issue there is: What is the legal framework under which you negotiate to get access?"

## Obstacles to Interaction

In the late summer of 2008, NCI's CPTC convened an international summit in Amsterdam to discuss data-sharing challenges and solutions. (See sidebar "Outcomes from Amsterdam.") Although that group focused on proteins, the challenges apply to sharing almost any sort of biotechnology data. According to Rodriguez, data sharing faces three categories of challenges: technology, infrastructure, and policy. "Moreover, each of those impacts the other two," says Rodriguez. So the challenges can be described individually, but they interact in practice.

The technology challenge in molecular biotechnology consists of several pieces. First, in genomics, proteomics and other fields, researchers use a range of technologies, such as mass spectrometry, tandem mass spectrometry, liquid chromatography and so on. That makes a variety of data that must somehow be compared. Worse still, the same instrument used in two labs can create different results just because the instrument gets calibrated in different ways. "So data from the same kind of instrument used with the same reagents but in two different labs can pump out data that are not comparable," Rodriguez says. The next technological challenge comes from the "flavor" of data being used—raw or processed. The raw data is just like it sounds, uncooked, not processed in any way, or as little as possible. If the data are processed, different data sets can only be compared when the exact processing can be taken into account, and the data must be

adjusted accordingly. Even if a researcher can get raw data from an instrument, that device could put the information in a proprietary format that is incomprehensible to other devices or analysis packages. And that analysis makes up the last technological challenge in data sharing. "Researchers use multiple computational tools—the algorithms that extract knowledge," says Rodriguez. Those algorithms pull out relationships that might be missed otherwise, but it proves difficult to compare data that were analyzed in different ways.

To get at the infrastructure behind data sharing, imagine a transportation analogy: Cars, trucks, trains, jets, ships and so on make up the data; and garages, roadways, waterways, skies and such make up the infrastructure. So the infrastructure determines where the data can be stored and the paths that data can take from one spot to another. "No international or centralized network has emerged," Rodriguez says. "Since the ones available use their own fixed formats, researchers cannot gather information from all of the sites." He adds, "Today's repositories are a benefit, but it will remain problematic if they are not interoperable."

---

**"Information  
should be shared  
on a broad scale.  
It is invaluable."**

---

Policy makes up the last category of data-sharing challenges. "In terms of proteomics," Rodriguez says, "the challenge here is really fundamental. It will be responsible for establishing and ultimately enforcing the guidelines for the proteomics community, including the requirements for submitting data and the metrics that will be used to determine the quality of the data." For example, standards should require researchers to

provide the metadata that explain the details behind the experiments that produced the actual data.

For data related specifically to healthcare, other policy considerations also arise. Patrick L. Taylor—deputy general counsel and chief counsel for research affairs at Children's Hospital Boston and assistant clinical professor at Harvard Medical School—writes often about data sharing, and he sees several obstacles, such as avoiding misuse of the data and creating a level playing field that takes into account the goals of commercial interests and patients. He says, "Managing access to data and its uses in ways that respect people's privacy but meets everyone's goals in public health is a real challenge."

So far, though, Taylor thinks that companies could do a better job of sharing data. "Huge amounts of tissue and data get collected in clinical trials," he says, "but that just gets banked away, used in for-profit directives, even though the research subjects just volunteered." Despite that company-patient imbalance, Taylor adds, "I don't want to demonize companies. They operate in their own environment."

Moreover, Taylor does not encourage a forced approach to data sharing. Instead he wants to find ways that encourage data sharing and help everyone along the way. "We could create data pools and give companies some level of access in exchange for sharing some of their own data." With such data pools, Taylor says, multiple companies might not need to reproduce the same data, which they do today.

## **Putting Sharing to Work**

In some areas, technology already makes data sharing possible. (See sidebar "Putting Patients Together.") One example is the cancer biomedical informatics grid, or caBIG, which was started by NCI and still run by it. Buetow describes caBIG as the "information technology framework that supports 21st-century biomedicine." He adds, "It's a way that we can interconnect the entire biomedical

enterprise using current-art information technology." So caBIG takes available technology and uses it to connect basic researchers, biomedical scientists, physicians and anyone else interested in cancer—and, actually, healthcare in general.

Just as others have to meet the challenges of sharing proteomics data, caBIG developers needed to make it possible for users to access a range of data types and to make sense of how they interact. Much of the problem revolves around translation—finding ways that software can unravel all of the medical community's vocabularies. To do this, caBIG provides a range of web services that are designed to work with anything that connects to caBIG. "A key component of caBIG is interoperability," Buetow says. "We are technologically neutral. Information can come from Oracle, a MySQL database and others." For example, caAdapter can be mounted on top of a data resource to make that information available on the caBIG framework.

Virtually anyone around the world can use the caBIG technology. Some international biotechnology operations are already underway. For instance NCI formed a partnership with Duke University related to international clinical trials. "So Duke established a partnership with the Beijing Cancer Hospital to get participation with Chinese colleagues," Buetow says. "They are using caBIG so that a trial being run in Durham, North Carolina, can recruit participants in Beijing, China."

In addition NCI developed a partnership with the Institute of Cancer Research in London. "They are installing a framework called Onyx that will be interoperable with caBIG. So we can interconnect between the U.K. and the U.S."

Despite being called a cancer grid, caBIG goes beyond cancer. "There is nothing cancer-specific about it," Buetow says. Instead NCI scientists hope that this system can draw together a range of health professionals around

the world. "In developing countries in particular," Buetow says, "this technology could help scientists become part of a bigger framework. These scientists could contribute their expertise to the field without building all of the components required in biotechnology or biomedicine."

## Speeding up Data Transmission

Many of the desired applications of data sharing, though, still hit information bottlenecks. As Feng and his colleagues found with BLAST, sequence searches could run faster by adding the parallel-computing capabilities of mpiBLAST. But even mpiBLAST is not always enough.

Sequence searches—even when done fast—still produce large amounts of data, which are not easy to move. So Feng and Pavan Balaji of the Argonne National Laboratory worked with some colleagues to develop ParaMEDIC, which stands for: parallel metadata environment for distributed I/O and computing. In fact, I/O—the input/output, or simply getting information into and out of computing resources—can really slow down data sharing.

To get around that, Feng and Balaji use ParaMEDIC to turn the original data into a code. With sequences, for example, ParaMEDIC uses GenBank Identifiers, which represent sequence strings. So instead of needing to grab a long length of bases—cytosine, guanine, thymine, cytosine and so on—ParaMEDIC just uses an identifier.

To see how well ParaMEDIC could really work, Feng and Balaji took on a tough problem. Scientists at the Virginia Bioinformatics Institute at Virginia Tech wanted to find the missing genes in 567 genomes from microbes, which required  $2.63 \times 10^{14}$  sequence searches. To do those searches, Feng and Balaji created a team of researchers, plus eight supercomputers scattered across the United States. The results consisted of 0.97 petabytes of data—almost a quadrillion bytes. To add the I/O side, they

planned to send the results—by Ethernet—to Tokyo. Sending the data in the conventional way would have taken about three years. With ParaMEDIC, the super computers cranked out the sequence searches, then crunched the results into a GenBank Identifier code. That crunching step turned the 0.97 petabytes into about four gigabytes, or reduced the data by roughly 250,000 times. As a result, the Feng and Balaji team computed the missing genes, sent the information from the United States to Japan, and had computers in Japan turn the code back into the original data—all in just 10 days.

This application could find lots of uses in biotechnology. "Say that you are a pharmaceutical company that has petabytes of sequence-search data stored around the world and you need to bring it to one place for some reason—back-up store or large-scale experiment," Feng says. "ParaMEDIC will enable the information to be shipped and reconstituted in a fraction of the time that it would take to recompute all the information locally."

In general, sharing data will remain under development, probably indefinitely. New research tools and growing data pools will require ongoing technological advances to keep the sharing doable. With every advance, though, sharing data will increase around the world.

## Outcomes From Amsterdam

When proteomic experts gathered in Amsterdam on August 14, 2008, to attend the International Summit on Proteomics Data Release and Sharing Policy, they focused on ways to get proteomic data into the public domain. "Our primary focus was on policy," says Henry Rodriguez, director of clinical proteomic technologies for cancer at the U.S. National Cancer Institute.

One policy decision involved when data should be released. In Amsterdam, Rodriguez and his colleagues concluded that it depends on the source of the data. If the data come

from an individual researcher's lab, the data should be released when the work gets published. For large-scale community projects designed to advance science in general, however, the data should be released as they are generated, provided that appropriate procedures exist to control the data quality.

In Amsterdam, the experts also considered what kind of data should be made available. "Raw data are the data that should go into the public domain," Rodriguez says. "Even if you agree to release raw data, though, they must be extremely well annotated with metadata. That defines the quality of the data itself."

Although many details must still be resolved, the intent is certain. "It is clear to me and others," Rodriguez says, "that data sharing expands and expedites research findings, especially where they are applicable to disease."

## Putting Patients Together

In 2004 a trio of M.I.T engineers—brothers Ben and Jamie Heywood and long-time friend Jeff Cole—founded PatientsLikeMe. In fact, this project really started in 1998, when another Heywood brother, Stephen, was diagnosed with amyotrophic lateral sclerosis (ALS), often called Lou Gehrig's disease. Although ALS is always fatal, slowly destroying the central nervous system, the Heywoods started looking for ways to give Stephen the best life that he could have. In 1999 Jamie founded the ALS Therapy Development Institute to speed up the generation of new treatments. Beyond finding new molecules, though, the Heywoods and Cole wanted to do even more. As described on the PatientsLikeMe website: "Our goal is to enable people to share information that can improve the lives of patients diagnosed with life-changing diseases. To make this happen, we've created a platform for collecting and sharing real world, outcome-based patient data ([patientslikeme.com](http://patientslikeme.com)) and are establishing data-sharing partnerships with doctors, pharmaceutical and medical device

companies, research organizations and non-profits."

This work goes beyond ALS. In fact, PatientsLikeMe plans to soon cover more than 50 diseases. It already provides communities for people with depression, HIV/AIDS, multiple sclerosis, Parkinson's disease and other afflictions. Perhaps most important of all, PatientsLikeMe reveals some of the breadth behind the ways that people can collect and distribute information.

"PatientsLikeMe consists of groups of people

coming together to share data in dramatically new ways," says Patrick L. Taylor, deputy general counsel and chief counsel for research affairs at Children's Hospital Boston and assistant clinical professor at Harvard Medical School, and a well-known expert on data sharing. "These data become a source of further data sharing. It's patient-specific, phenotypically interesting, longitudinal data shared by patients themselves."

© 2009 Scientific American, Inc. and Worldview, All Rights Reserved. Reproduction in whole or in part without permission is prohibited.

# Expert Opinion

1. The challenge of proteomics
2. Overcoming barriers
3. Accelerating clinical translation
4. Final thoughts
5. Expert opinion

## Proteomics: a long and winding road to medical diagnostics

Henry Rodriguez

National Cancer Institute, Clinical Proteomic Technologies for Cancer, Office of Technology and Industrial Relations, 31 Center Drive, MS 2590, Bethesda, MD 20892, USA

**Background:** The addition of protein biomarker panels to the cancer diagnostic armamentarium is an area of considerable interest in medicine. The discovery that proteins and peptides are 'leaked' by tumors into clinically accessible bodily fluids such as blood and urine has led to the possibility of diagnosing cancer at an early stage or monitoring response to treatment simply by collecting these fluids and testing for the presence of cancer-related biomarkers. Prostate-specific antigen and cancer antigen-125 are examples of blood-borne cancer protein biomarkers that are now being used in the clinic. However, the measurement of individual biomarkers has clinical limitations with respect to both sensitivity and specificity. For this reason, combinations of protein/peptide analytes are under intense investigation, as biomarker panels can potentially bring greater sensitivity and specificity to cancer screening than any one analyte alone. **Objective:** This perspective article highlights some of the challenges that hinder the translation of proteomics to clinical diagnostics. **Conclusion:** Tremendous advances have been made in genomics over the past decade, and this field holds great promise for understanding, treating and even preventing many forms of cancer; although genes are the 'recipes' of the cell, proteins are the critical molecular players that drive both normal and disease physiology. If the potential of personalized medicine is to be realized, it needs to include this next generation of molecular diagnostics.

**Keywords:** bioinformatics, biomarkers, biospecimens, cancer, clinical proteomics, data sharing, genomics, mass spectrometry, pipeline, proteomic technologies, proteomics, shotgun proteomics

*Expert Opin. Med. Diagn.* (2009) 3(3):219-225

### 1. The challenge of proteomics

With well over 1000 disease-associated protein biomarker candidates described in the scientific literature [1], one would expect the diagnostic market to be flooded with new clinical tests, but this is not the case. The sobering reality is that very few of these biomarker candidates have been validated, and even fewer have made it into a medical diagnostic product [2]. This discrepancy between discovered candidates and approved protein diagnostic markers indicates an issue within the candidate biomarker pipeline.

#### 1.1 Biological complexity

The biggest conceptual challenge inherent in proteomics lies in the proteome's increased degree of complexity compared with the genome [3]. For example:

- One gene can encode more than one protein. The human genome contains ~ 21,000 protein-encoding genes, but the total number of proteins in human cells can be in the millions owing to genomic splice variants.
- Proteins are continually being modified during and post-production. As a result, proteins can vary considerably from one person to another, under different

environmental conditions, or even within the same person at different ages or states of health.

- Proteins exist in a wide range of concentrations in the body. For example, the concentration of albumin in blood is more than a billion times greater than that of interleukin-6. This dynamic range makes it extremely difficult to find proteins of low abundance in a mixture, which is widely believed to contain the most important proteins for cancer (Figure 1).

The challenge for protein scientists is to navigate through this enormous complexity in order to identify true biological changes that are associated with disease. The technologies for overcoming this challenge are developing, but the community needs to work out issues of variability and reproducibility before clinical proteomics can really take off.

### **1.2 Experimental variability**

Proteomic technologies can help make sense of the complex proteome and hold great promise for the discovery of new cancer biomarkers. In recent years, however, studies that have applied proteomic technologies – such as mass spectrometry and affinity-based detection methods – to clinical applications have met with some disappointment [4]. Part of the problem lies with the variability that exists in every aspect of the proteomics pipeline (experimental design, bio-specimen collection and preparation, protein/peptide fractionation and detection, and protein/peptide identification and quantification – data analysis) (Figure 2) [5]. In addition, the variety of platforms used combined with individual labs developing their own procedures for equipment calibration has led to pervasive problems in the comparison of research results from two different labs, as each lab calibrates its equipment differently. Also, no universal standardized sources exist for experimental reagents, adding to the difficulties of accurately comparing or replicating data generated across laboratories. All of these analytical and pre-analytical variables pose significant challenges to the translation of discoveries to clinical applications by introducing experimental variability on top of biological complexity.

'Shotgun' proteomics is the most commonly used approach for the discovery and subsequent identification of proteins in complex mixtures (e.g., plasma). Shotgun proteomics requires several complex steps – from experimental design to data analysis – each of which can introduce variability. In fact, when 5 separate proteomic laboratories were asked to analyze a basic sample containing only 20 proteins using this approach, 5 completely different data sets were returned [5]. To put this into perspective, a human blood sample contains at least 100,000 different proteins. If laboratories cannot reach the same conclusion analyzing a common sample, it cannot be surprising that many protein biomarkers fail to be clinically validated.

### **1.3 Technical limitations**

Protein candidate biomarker discovery typically involves attempting to identify proteins present in very low abundance.

Although there can be a significant division of opinion over whether the low abundance plasma components are more 'interesting' or clinically meaningful than those of high abundance (or vice versa), everyone can agree that plasma is believed to be rich in biomarker candidates. However, today we know that blood poses significant challenges owing to the fact that only 12 proteins comprise up to 96% of the protein mass in plasma (Figure 3), and the total abundance of proteins covers a logarithmic scale spanning almost 12 orders of magnitude [6].

The Human Proteome Organization (HUPO) published a report in 2006 on a collaborative study involving 18 laboratories to characterize a common sample of human plasma [7]. Although this study showed that 900 proteins were identified with high confidence (> 95%), it was the abundant proteins that had very high and reproducible representation in the data. Proteins in the ranges where most believe biomarker candidates are located – those of low abundance – were poorly represented.

## **2. Overcoming barriers**

To make clinical cancer proteomics a reality, the scientific community needs to invest in much needed technologies and infrastructure in order to address the issues of reproducibility that plague candidate biomarker discovery. This means eliminating experimental variability as much as possible so 'true' biology can be discovered.

### **2.1 Optimize proteomic technologies and develop appropriate standards**

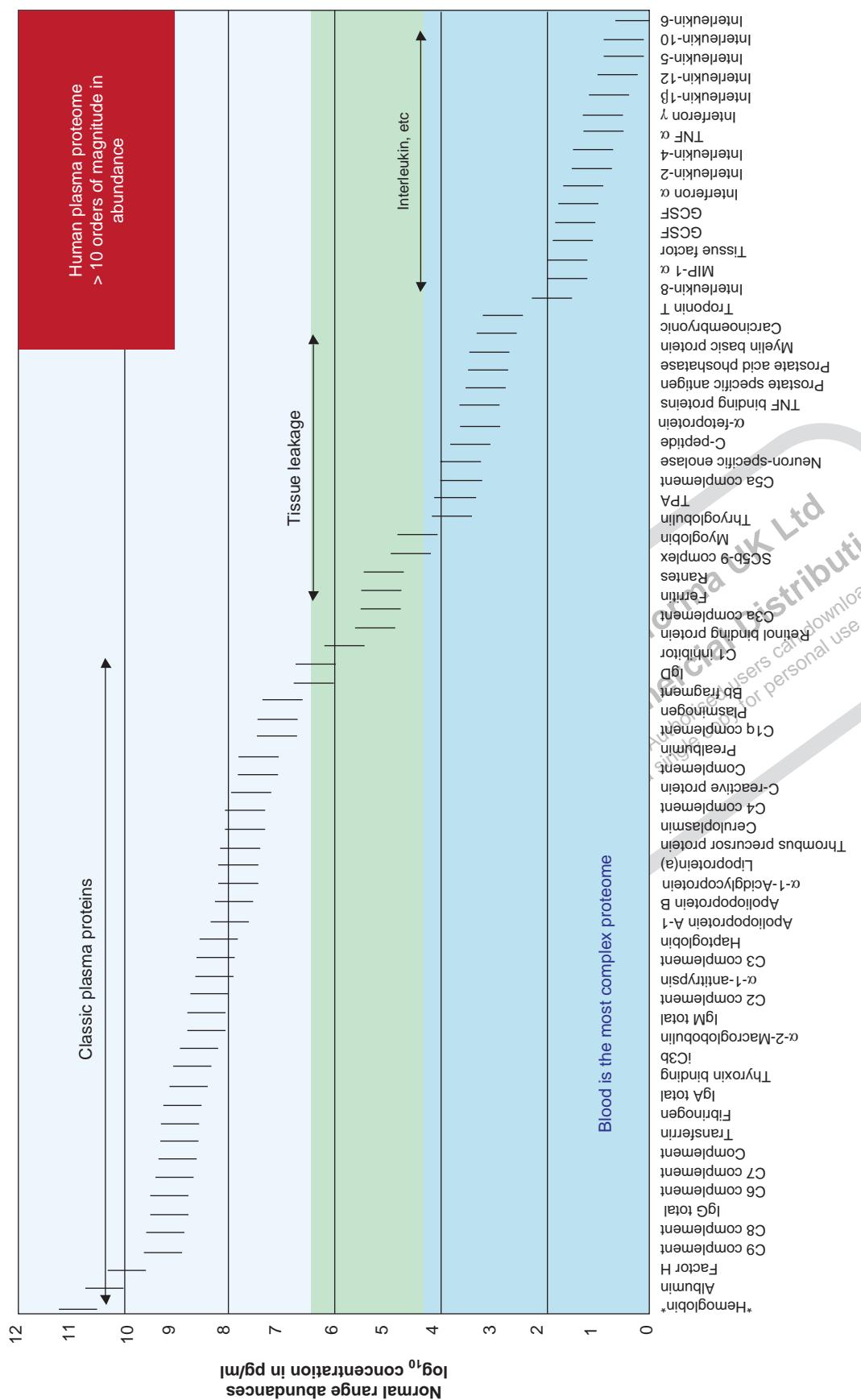
Current and emerging protein measurement technologies need to be optimized and calibrated through the use of standard protocols and performance reagents to produce comparable results between different laboratories.

### **2.2 Standardize procedures for collecting, processing and storing biological samples used in proteomics research**

The use of high-quality biospecimens is critically important for proteomic research because the output – the data – is only as good as the input. The methods of biological sample preparation need to be made more consistent to reduce variability in experimental results. Uniform sample quality, as well as access to large numbers of high-quality samples, will lead to more reliable results.

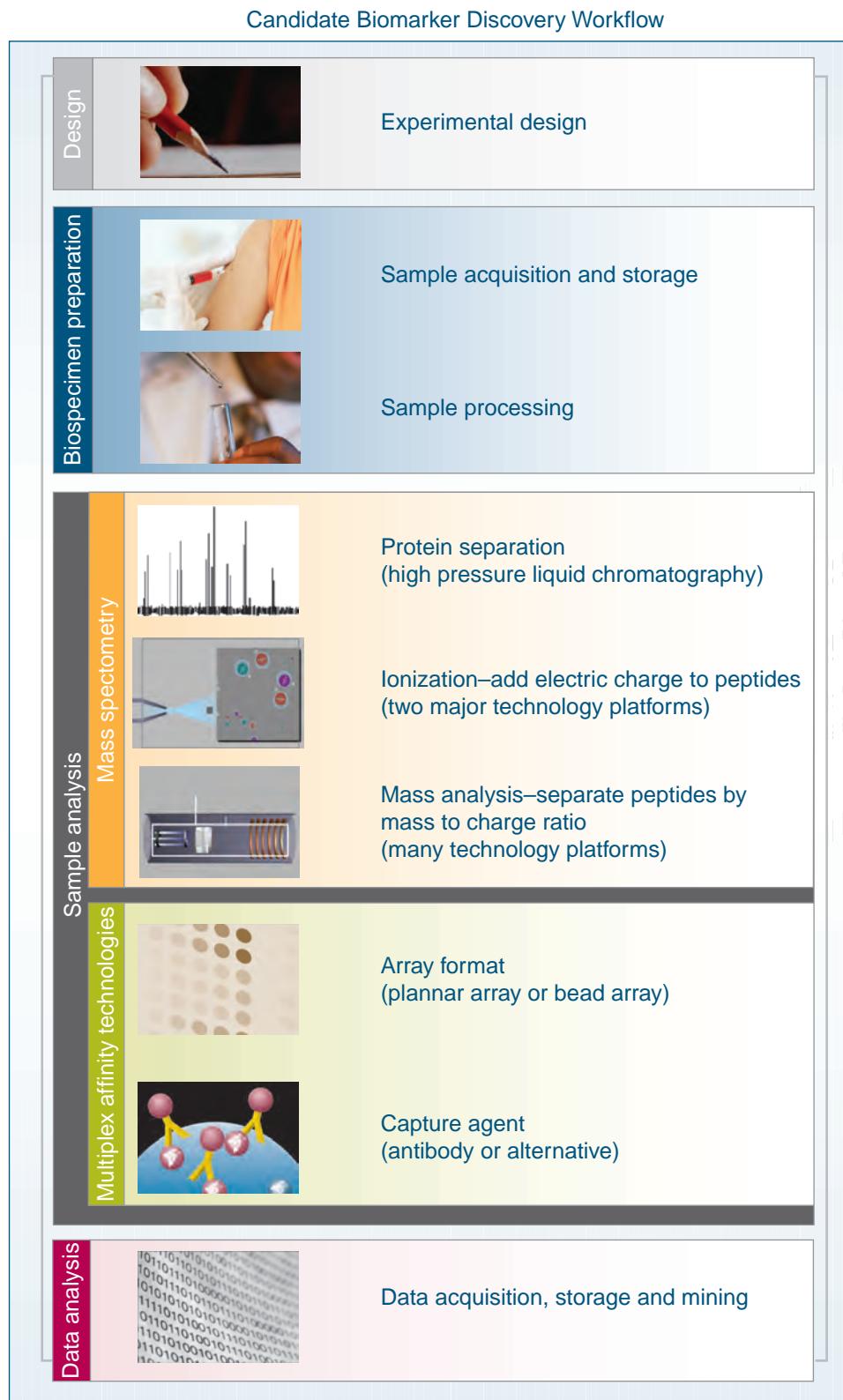
### **2.3 Make high-quality reagents available and accessible**

In particular, capture reagents (e.g., antibodies) that can be used in protein microarrays, as well as other techniques used to measure proteins, are an essential resource. However, a key challenge for proteomic researchers is acquiring high-quality, well-characterized monoclonal antibodies [8]. Although numerous commercial reagent suppliers make antibodies available for research, they tend to be expensive and may or may not be extensively characterized.

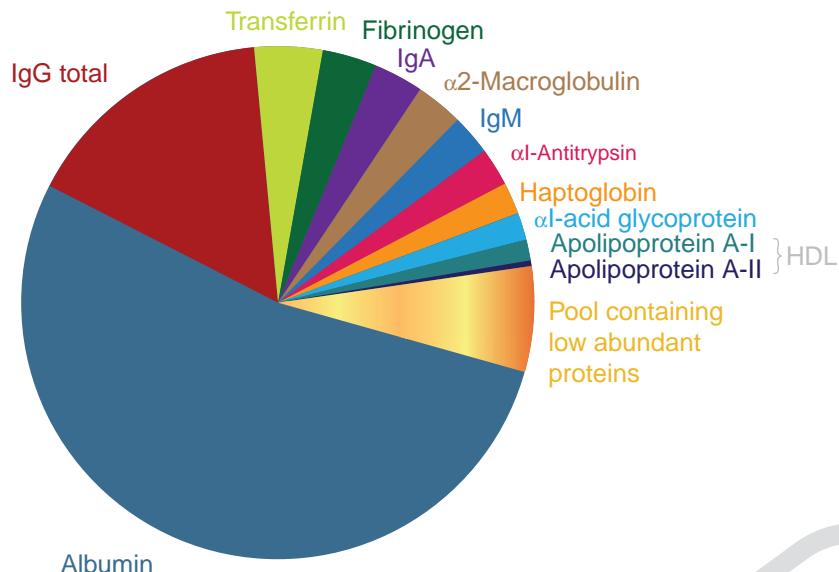


**Figure 1. Abundance of 70 Reference Protein Analytes in Plasma.** The dynamic range of proteins in human plasma spans almost 12 orders of magnitude. It is believed that many protein biomarker candidates will be found in the low abundance range, which poses a significant challenge for discovery.

Adapted from Anderson, N.L., and Anderson, N.G. (2002) The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular Proteomics*, 1, 845-867.



**Figure 2. Candidate Biomarker Discovery Workflow.** Each step within this workflow contributes to experimental variability. Such variability poses a significant challenge to the translation of proteomic discoveries to clinical applications.



**Figure 3. Twelve Proteins Comprise up to 96% of the Protein Mass in Plasma.** Developing ultra-sensitive detection methods that have a large dynamic range has been a significant challenge and identifying dilute proteins from a complex mixture has proven to be especially difficult.

#### 2.4 Develop technologies that can quantify proteins across the entire concentration range as well as detect modified versions of proteins

Enormous variations in protein concentrations and modifications are found within cells and fluids, and developing ultra-sensitive detection methods that have a large dynamic range has been a significant challenge. Capturing and identifying dilute proteins from a complex mixture has proved to be especially difficult. In addition, many of the best techniques for finding low-abundance proteins are not quantitative – that is, they can determine whether the protein is present or not, but they cannot measure its concentration.

#### 2.5 Develop common bioinformatics resources with shared algorithms and standards for processing, analyzing and storing proteomic data

Proteomic informatics tools that permit data sharing and computation among laboratories are essential for rapid progress in the field.

#### 2.6 Include a verification step in the candidate protein biomarker pipeline

Protein biomarker candidates should be ‘pre-qualified’ before costly clinical trials. Pre-qualification can be accomplished by implementing a verification step in the biomarker pipeline. Independent research labs verify that the analytical measurements are real, providing assurance that the potential biomarkers are a strong candidate for advancement to the next stage of development-clinical validation.

#### 2.7 Adopt an interdisciplinary team approach to science

No single laboratory working on its own could possibly examine all of the candidate biomarkers, develop all of the necessary technologies, or assemble all of the pieces of evidence required to understand the molecular mechanisms of disease. It will require many laboratories working together to accomplish these goals.

In many ways, the challenges facing the clinical proteomics community are comparable with those that were faced by the genomics community before the Human Genome Project – the current technology enables the sampling of only a small portion of the proteome, at different levels of quality. Improvements made in DNA sequencing along with standardized protocols and methodologies were responsible for the success of the Human Genome Project, and each team was able to collect and analyze all data assembled at the end of the project as a result. Only when this was made possible did the genomics revolution take off.

### 3. Accelerating clinical translation

The US National Cancer Institute (NCI) recognizes the promise of clinical proteomics for the early detection and treatment of cancer and also acknowledges the technical and organizational challenges that face the proteomics community. Responding to the needs of the community, the NCI launched the Clinical Proteomic Technologies for Cancer (CPTC) initiative in 2006 as a way to bring together the best minds in proteomics – to create an entire community that is devoted to

fixing the protein biomarker pipeline – because this is far too great an endeavor for a single institution.

The goal of the CPTC initiative is to address and possibly reduce the layers of variability at every step of the biomarker discovery pipeline, from experimental design to data analysis, in order to accelerate the translation of protein biomarker discoveries to clinical utility. Investigators will then have the assurance that protein/peptide measurement results are due to changes in a biological sample and not to variability in the instrument, assay performance, reagents, operator, or site. Ultimately, this process will significantly improve the quality of biomarker candidates that enter the clinic.

To help meet this ambitious goal, CPTC has joined with scientists from nearly 50 federal, academic and private-sector organizations to compose three major, integrated programs: the Clinical Proteomic Technology Assessment for Cancer (CPTAC) program, the Advanced Proteomic Platforms and Computational Sciences program, and the Proteomic Reagents and Resources Core. CPTC has recently launched a Reagents Data Portal through which highly characterized monoclonal antibodies to human proteins associated with cancer are available to the scientific community. The data portal can be accessed through the CPTC website [9].

#### **4. Final thoughts**

The standards that are needed to conduct large-scale genomic studies are also needed for clinical proteomics – standards for collecting and processing clinical samples, conducting experiments, and collecting and analyzing data – so that teams of scientists in different laboratories can collate and analyze their data to achieve meaningful results. Developing these standards is what the CPTC initiative strives to accomplish – it is building the highway from discovery to the clinic. When this is achieved, the proteomics revolution will begin.

#### **5. Expert opinion**

Visionaries brought the Human Genome Project to life, but it was improvements made in DNA sequencing technologies

that made this endeavor possible. The technologies became high-throughput, reliable and reproducible. Had the project moved forward without these much needed technological improvements, the genomics community would never have been able to live up to its promise, and the opportunities afforded by the Human Genome Project would never have been realized. Proteomics stands at a similar juncture. Do we improve the technologies and methodologies first – do we build the highway? Or do we blaze ahead and hope for the best?

The field of clinical proteomics is viewed with skepticism largely because we have not yet delivered on what we promised our efforts could provide: clinically validated disease-specific biomarkers. Although the lack of success so far has been disappointing, it is important to realize that the science of proteomics is still young, and we are starting to acknowledge the roadblocks that need to be addressed. If we are going to fix the reputation of the field, it is imperative that the issues of reproducibility, which plague the proteomics community, first be addressed. The challenge now is to avoid overstating our current capabilities, take a step back, and strengthen our weaknesses as a field. The development of standard methodologies and reagents – from experimental design to data analysis – will ensure reproducibility between laboratories, resulting in a portfolio of biomarkers that can be trusted to enter the clinic. We have a lot to prove.

A Human Proteome Project is now under discussion by a small group of scientists as a natural successor to the Human Genome Project [10]. There is no doubt that such an undertaking will bring enormous benefit to patients in the form of personalized medicine, but before we commit to such an endeavor and risk credibility of the field even further, we must be cautious and understand what made the Human Genome Project successful – they built the highway first.

#### **Declaration of interest**

The author states no conflict of interest and has received no payment in preparation of this manuscript.

## Bibliography

1. Polanski M, Anderson NL. A list of candidate cancer biomarkers for targeted proteomics. *Biomarker Insights* 2006;2:1-48
2. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Cancer* 2005;5:845-56
3. Seet BT, Dikic I, Zhou MM, Pawson T. Reading protein modifications with interaction domains. *Nat Rev Mol Cell Biol* 2006;7:473-83
4. Check E. Proteomics and cancer: Running before we can walk? *Nat* 2004;429:496-7
5. Clinical proteomic technologies for cancer. 2007 annual report: building the foundation for clinical cancer proteomics.
- Available from: <http://proteomics.cancer.gov/library/reports.asp>. [Accessed 18 November 2008]
6. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 2002;1:845-67
7. States DJ, Omenn GS, Blackwell TW, et al. Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* 2006;24:333-8
8. Haab BB, Paulovich AG, Anderson NL, et al. A reagent resource to identify proteins and peptides of interest for the cancer community: A workshop report. *Mol Cell Proteomics* 2006;5(10):1996-7. Epub 24 Jul 2006
9. Clinical proteomic technologies for cancer Available from: <http://cancer.proteomics.gov>
10. Pearson H. Biologists initiate plan to map human proteome. *Nat* 2008;452:920-1

## Affiliation

Henry Rodriguez PhD MBA

<sup>†</sup>Author for correspondence  
Director Clinical Proteomic Technologies for Cancer,  
Office of the Director,  
Office of Technology and Industrial Relations,  
National Cancer Institute,  
31 Center Drive, MS 2590,  
Bethesda, MD 20892, USA  
Tel: +1 301 496 1550;  
E-mail: rodriguezh@mail.nih.gov

Copyright © 2009 Informa UK Ltd  
Not for Sale or Commercial Distribution  
Unauthorised use prohibited. Authorised users can download, display, view and print a single copy for personal use.

# Commentary: Complementary fields of proteomics and genomics are cornerstone of personalized medicine

BY DR. HENRY RODRIGUEZ

**T**ODAY, THE WORLD OF personalized medicine is very gene-centric. This is not surprising given that genomic studies are among the most promising strategies to help advance personalized medicine efforts, particularly through the identification of disease susceptibility genes and disease-specific gene expression profiles. While genomics will always remain a cornerstone of personalized medicine, these studies alone cannot capture the complete view of disease processes.

While genes are the "recipes" of the cell, containing all of the instructions for assembly, proteins are the products of these recipes. Proteins function as the cellular "engines"—the molecules that do much of the actual work to keep the cell and the body functioning—and also drive disease physiology. So while genomics may provide the propensity of developing a certain disease, proteins may diagnose what is happening in a patient in real-time, such as whether a patient has a disease.

Biology is complex, and the translation from gene to protein is not a clear-cut process. This means that gene expression profiles do not accurately represent what is happening at the protein level because



**Dr. Henry Rodriguez,**  
Director, Clinical  
Proteomic Technologies  
for Cancer, Office of the  
Director, National Cancer  
Institute

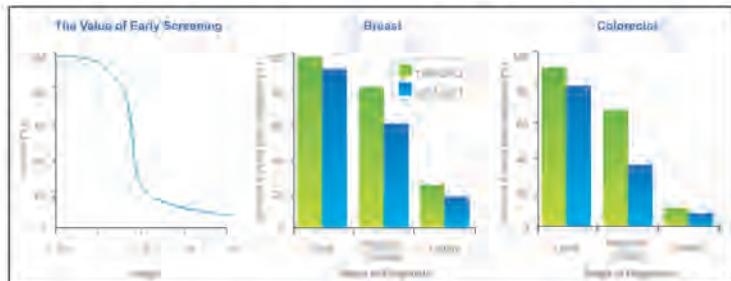
personalized medicine.

There is a high unmet clinical need for early disease detection, especially in the area of cancer. It is well known that early detection extends survival and, if diagnosed early enough, can perhaps cure the disease (see graphic). As such, the need for relevant and valid biomarkers is key to detecting, treating and even preventing cancer and other diseases. In fact, cancer is one of the areas that can benefit greatly from personalized medicine and where we have first seen direct results.

Biomarkers are molecules such as proteins that are present in tissue or fluids. The finding that tumors "leak" proteins into blood and urine has led to the possibility of detecting cancer at a very early stage by collecting these bodily fluids and testing for the presence of cancer-related biomarkers. Protein biomarkers also hold promise for monitoring the response to cancer therapy or detecting tumor recurrence following treatment.

Protein biomarkers currently exist in the clinic. For example, protein specific antigen (PSA) and cancer antigen 125 (CA-125) are measured to detect for the presence of prostate and ovarian cancer, respectively. However, individual biomarkers have limitations with respect to both sensitivity and specificity. The goal within the clinical proteomics community is to provide panels of protein biomarkers that will ultimately provide much greater sensitivity and specificity than individual analytes.

The cancer proteomics community does not feel a shortage of candidate protein biomarkers. As of 2006, there were over 1,200



that the field is still young. Only now are we starting to recognize and address the barriers to progress, which include both technology and systems barriers. The technology barriers include: an inability to reproduce experimental data; a large dynamic range of proteins that need to be measured; limited interoperability across instruments and platforms; the number of platforms available; a lack of multiplexed detection platforms suitable for clinical use; and insufficient informatics tools for data capture and analysis.

The systems barriers identified include: a lack of high-quality reagents (e.g., monoclonal antibodies); an inadequate supply of high-quality biospecimens; a lack of standards and protocols; and cultural barriers to multidisciplinary research.

All of these variables together pose significant challenges to the translation of biomarker discoveries into clinical applications. In order to make clinical proteomics a reality, the community must first invest in much-needed technologies and infrastructure to address these critical issues in order to advance the field.

Recognizing the promise of clinical proteomics for the early detection and treatment of cancer—and recognizing that action is needed—the National Cancer Institute has taken a lead role in bringing the field to the next level by launching the Clinical Proteomic Technologies for Cancer (CPTC) initiative in 2006. This initiative is made up of the best minds in proteomics, including scientists from nearly 50 federal, academic and private-sector organizations, that are working together to fix the protein biomarker pipeline—far too great an endeavor for a single institution.

Through the development of standard methodologies and reagents throughout the entire biomarker pipeline, CPTC aims to provide investigators with the assurance that protein/peptide measurement data are based on biology rather than instrumentation, assay performance, reagents, operator, or site—all of which can have variability. This will accelerate the translation of protein biomarker discoveries to the clinic.

In fact, CPTC has already launched a Reagents Data Portal through which highly characterized monoclonal antibodies to human proteins associated with cancer are available to the greater scientific community. The data portal can be accessed through the CPTC Web site at:

<http://proteomics.cancer.gov>.

In many ways, the field of proteomics can learn from its technologically more advanced counterpart, genomics. Unraveling genomic complexity was directly tied to improvements made in technologies and method-

ologies, which resulted in high-throughput, reliable and reproducible results. The rapid uptake of shared technologies and methodologies resulted in reduced cost, increased efficiencies, and also made the sequencing of the Human Genome Project successful. Microarrays have since allowed for characterization and quantitation of gene expression and chip technologies are significantly increasing in density and sensitivity, allowing whole genome scans on one or two chips.

The standards that were employed to advance the field of genomics to where it is today, permitting large-scale genomic studies, are the same standards that are needed for clinical proteomics. We need standard methodologies and reagents at every step along the biomarker pipeline, so that investigators in different proteomics laboratories can generate reproducible data to achieve meaningful results. It is the goal of the CPTC initiative to develop such standards so proteomic technologies will also become high-throughput, reliable and reproducible. Only when this is accomplished will the next generation of molecular diagnostics become available for the early detection of cancer and treatment. Genomics and proteomics are the cornerstone of personalized medicine. Together, they will make great strides in patient care. **DDN**

**WHILE GENOMICS WILL ALWAYS REMAIN a cornerstone of personalized medicine, these studies alone cannot capture the complete view of disease processes. While genes are the "recipes" of the cell, containing all of the instructions for assembly, proteins are the products of these recipes. Proteins function as the cellular "engines"—the molecules that do much of the actual work to keep the cell and the body functioning—and also drive disease physiology. So while genomics may provide the propensity of developing a certain disease, proteins may diagnose what is happening in a patient in real-time, such as whether a patient has a disease."**

a direct correlation does not exist between the levels of gene transcripts and resulting protein expression. Additionally, proteins undergo post-translational modifications (PTM), including phosphorylation, glycosylation, lipidation and cleavage. PTMs and protein-protein interactions are also known to play a significant role in disease processes and could serve as both disease-specific biomarkers and therapeutic targets.

Together, these complementary fields, genomics and proteomics, are absolutely necessary for understanding the molecular underpinnings of disease and for enabling

candidate protein biomarkers described in the scientific literature. With such robust discovery efforts, one would expect the molecular diagnostics market to be flooded with new clinical tests but this is not the case. Unfortunately, most candidates have not been validated and in fact only about a dozen have actually been FDA-approved as tumor-associated antigens. Such a discrepancy between discovery and validation suggests that issues within the biomarker pipeline need to be addressed.

Although the lack of success to date has been disappointing, it is critical to remember

Dr. Henry Rodriguez is an internationally recognized expert in advanced molecular-based cancer technologies, specifically proteomics, genomics, celomics and bioinformatics. He currently serves as the director of the Clinical Proteomic Technologies for Cancer programs within the NCI's Office of Technology and Industrial Relations, in the Office of the Director, overseeing scientific proteomic technology programs such as the Clinical Proteomic Technologies for Cancer Initiative (CPTC) and the Mouse Proteomic Technologies Initiative (MPTI).

## YOUR THOUGHTS HERE

Every month, *Drug Discovery News* will feature the thoughts, opinions and important discoveries of our readers on this page. It's an opportunity for you to sound off on the hot topic on everyone's tongues or to shine a light on an area you feel deserves more attention.

We will consider previously published pieces, though our preference is to publish original columns written for us and geared to our readers: the 30,000 lab managers, lead researchers and researchers who read us every month. Columns should be between 700 and 800 words and be accompanied by a picture of the author.

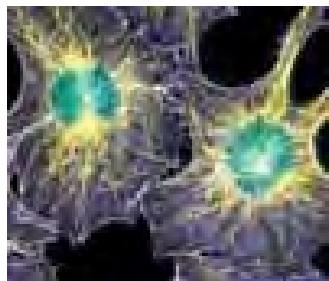
To discuss an idea for a guest commentary submission, contact Amy Swinderman, managing editor, at [swinderman@drugdiscoverynews.com](mailto:swinderman@drugdiscoverynews.com).

## CLINICAL PROTEOMICS HEADS INTO REAL WORLD

IMPROVED INSTRUMENTATION AND UNBIASED SAMPLES RENEW

PROMISE OF BIOMARKER PIPELINE

N. Leigh Anderson, Ph.D.



**Antibodies, a key component in protein diagnostics, are critical for demonstrating the presence and subcellular location of proteins.**

Given that proteins are the primary working parts of cells, it seems self-evident that proteomics should yield abundant clues to disease mechanisms, as well as numerous clinically useful biomarkers. Biomarkers in readily accessible bodily fluids such as plasma, in particular, offer the potential for rapid advances in patient care through early diagnosis, selection and monitoring of treatment, as well as acceleration of drug development.

It is therefore surprising that the rate at which new protein diagnostics have been approved by FDA over the past 12 years has steadily declined. In fact no protein biomarkers arising from proteomics appear to have entered broad clinical use so far. Why is this?

From the viewpoint of basic biology, biomarker development is probably as difficult as drug development. Both aim at discovery and verification of a disease-related biological invariant across an out bred human population, and they seem to have similar candidate attrition rates. Given the 100:1 revenue advantage of pharma compared to protein diagnostics, one is tempted to imagine that biomarker work should nevertheless be far less expensive and take less time than drug development. This wishful thinking is just that.

### SELDI Fiasco

Not so long ago (~2002), finding new cancer biomarkers in serum was made to look easy by applying an astonishingly simple new proteomics platform to a few samples from diseased patients and samples from a few healthy controls. This approach, commonly referred to as SELDI, combined three novel technology components, each of which is now known to be problematic.

The result was a general failure due to biases in the data (due in this case to machine drift between runs of cases and controls, but in other cases to sample processing and/or patient group selection). As a result, when analyses are repeated at other sites candidate disease patterns fail to replicate. Despite the efforts of dedicated biotech companies, and two of the largest clinical reference laboratories, SELDI tests for cancer have still not gained FDA approval.

While the reasons for this debacle are now well-understood and useful elements of the approach redeveloped in more rigorous form, clinical proteomics is only now recovering from the "SELDI bubble" caused by the initial excitement over this approach. Fortunately substantial parallel advances have been made (albeit with far less hype) in understanding critical sample requirements, in improving the performance of advanced MS (mass spectrometers) instrumentation, and in understanding the appropriate structure for a real biomarker pipeline.

It is difficult to overstate the importance of samples and experimental design in the operation of a biomarker pipeline. Two major factors arise: quality of the samples and number of samples.

High-quality samples are collected in such a way that there is no difference in collection or processing between groups: i.e., no bias. The typical number of samples required to convince diagnostic professionals that a biomarker is likely to have clinical utility (the Zolg number) is about 1,500, and technology platforms that cannot analyze this number of samples are not really usable in the later stages of biomarker verification.

### Enhanced Equipment

On the instrumentation front, two streams of technology have emerged as critical to the biomarker effort. On the one

hand, increasing resolution, sensitivity, and speed of high-end mass spectrometers now enables the detection of tens of thousands of tryptic peptides (and by inference thousands of proteins from which they came) in complex biological samples.

On the other hand, a separate stream of quantitative MS technology, measuring preselected peptide ions based on two mass parameters (parent peptide and a specific sequence fragment, in so called multiple-reaction monitoring, or MRM, mode) provides a capability to accurately (~10% CV) quantitate 100 or more peptides at much higher throughput.

The sensitivity of multiplex MRM technology can be extended down to the ng/mL level and below by abundant protein depletion combined with limited fractionation or specific capture of the target peptides on antipeptide antibodies (the SISCAPA technique), covering a majority of the known biomarker proteins detected in blood plasma.

It is now clear that a functional biomarker pipeline needs both of these approaches: shotgun methods to search large numbers of peptides and proteins for potential disease-related differences, albeit with a high false-discovery rate, and MRM methods to construct accurate high-throughput assays to be applied to relevant Zolg-scale sample sets to verify performance in real populations.

Evaluating, optimizing, and implementing these and other recent advances are critical to solving the general biomarker problem. To do so, however, requires enlarging the focus of our efforts from technology-centric academic proteomics to a multidisciplinary (though possibly virtual) biomarker pipeline.

### **Interrelated Efforts**

Productive relationships must be forged between disparate technology platforms and between technological, medical/biological, and statistical specialties. The **U.S. National Cancer Institute** is attempting to create a nucleus for this new approach in its Clinical Proteomic Technology Assessment for Cancer program within the broader Clinical Proteomic Technologies for Cancer initiative.

Beginning with a critical evaluation of existing technology and methods, the CPTAC teams have designed and carried out true multisite reproducibility studies of both approaches: shotgun unbiased discovery and targeted MRM assays. The results, recently presented and now submitted for publication, are revealing.

As has been expected based on earlier, less well-controlled studies (e.g., the HUPO plasma proteome exercise), the shotgun approaches produce a statistical samples of the peptides in the proteome under study and thus often show significant differences in the sets of peptides from run to run both within and between laboratories (with greater similarity at the protein level).

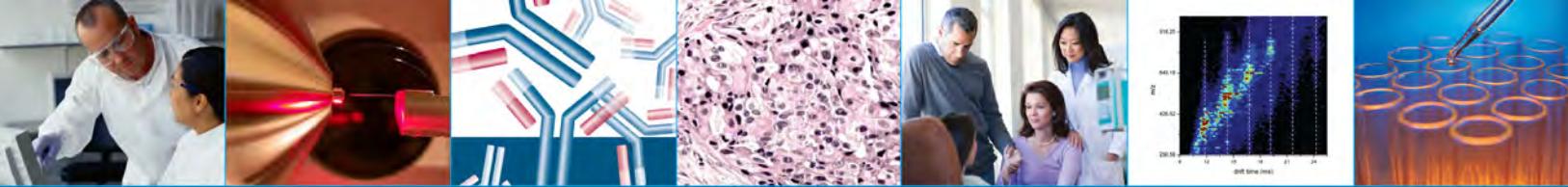
The need for replicate runs to approach asymptotic completeness in proteome coverage is thus an inherent statistical feature of the method. The targeted MRM assays, on the other hand, derived from a widely used accurate quantitation approach for small molecules, can yield results that are accurate, reproducible (in this case across eight sites), and of wide dynamic range provided we restrict attention to a set of up to several hundred prespecified peptides.

These studies have confirmed the roles and fitness of these two approaches for discovery and verification of candidates in the biomarker pipeline, and provide confidence that both can be practiced effectively in multiple laboratories.

In parallel, it appears that MS-based measurements can deliver high-quality results in clinical laboratories as well. A specific example highlighting these issues is the clinical assay for plasma thyroglobulin, a thyroid-specific protein used to detect recurrence of thyroid cancer in patients whose diseased thyroids have been removed.

Andrew Hoofnagle, M.D., Ph.D., recently demonstrated an MS-based SISCAPA assay for peptides from thyroglobulin designed to circumvent several well-known and high-prevalence interferences plaguing the existing commercial immunoassays for this protein.

The prospects for major progress in protein biomarkers in readily accessible bodily fluids thus appear considerably brighter than even a year or two ago. The clinical and economic value of early detection of diseases like cancer,



## 4.13 CPTC Publications

Addona T., Abbatiello S.E., Skates S. J., Bunk D. M., Schilling B., Spiegelman C. Zimmerman L. J., Ham A-J, Keshishian H., Hall S. C., Allen S., Anderson N. L., Blackman N. L., Borchers C., Buck C., Cardasis H. L., Cusack M. P., Dodder N. G., Gibson B. W., Held J. M., Hiltke T., Jackson A., Johansen E.B., Kinsinger C. R., Li J., Mani D.R., Mesri M., Neubert T.A., Niles R.K., Paulovich A.G., Pulsipher T.C., Rodriguez H., Rudnick P. A., Smith D., Tabb D. L., Tegeler T. J., Variyath A. M., Vega-Montoto L. J., Wahlander Å., Waldemarson S., Wang M., Whiteaker J. R., The NCI Clinical Proteomic Technology Assessment for Cancer (CPTAC) Network, Fisher S. J., Liebler D.C., Regnier F.E., Tempst P., Carr S.A. A Multi-site Assessment of Precision and Reproducibility of Multiple Reaction Monitoring-based Measurements: Toward Quantitative Protein Biomarker Verification in Human Plasma. *Nat Biotechnol.* 2009 Jul;27(7):633-41. Epub 2009 Jun 28.

Ahn N.G., Shabb J.S., Old, W.M., and Resing, K.A. (2007) Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chemical Biology*, 2:39-52.

Alves P., Arnold R. J., Clemmer D. E., Li Y., Reilly J. P., Sheng Q., Tang H., Xun Z., Zeng R., Radivojac P. (2008) Fast and accurate identification of semi-trypic peptides in shotgun proteomics. *Bioinformatics*, 24 (1): 102-109.

Anderson, N. Leigh; Anderson, Norman G.; Pearson, Terry W.; Borchers, Christoph H.; Paulovich, Amanda G.; Patterson, Scott D.; Gillette, Michael; Aebersold, Ruedi and Carr, Steven A. (2009) A Human Proteome Detection and Quantitation Project: hPDQ. *Molecular Cell. Proteomics*, Published online January 7, as Manuscript R800015-MCP200.

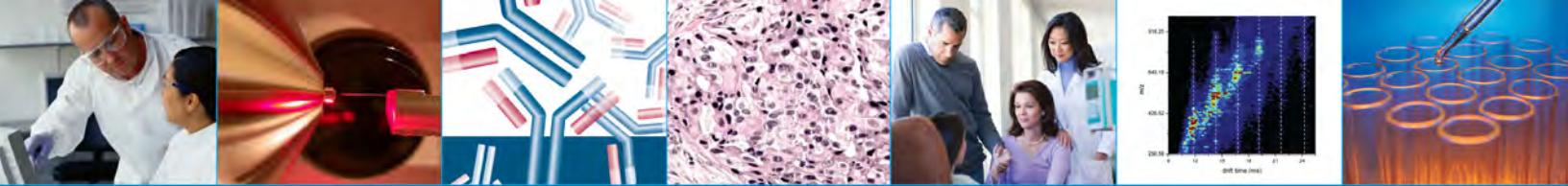
Anderson, N. Leigh; Jackson, Angela; Smith, Derek; Hardie, Darryl; Borchers, Christoph and Pearson, Terry W. (2009) SISCAPA Peptide Enrichment on Magnetic Beads Using an In-line Beadtrap Device. *Molecular Cellular Proteomics*, 8:995-1005.

Arnett D.R., Jennings J.L., Tabb D.L., Link A.J., Weil P.A. (2008) A proteomics analysis of yeast Mot1p protein-protein associations: insights into mechanism. *Mol Cell Proteomics*, 7, 2090-2106.

Baker E.S., Tang K., Danielson III W.F., Prior D.C., Smith R.D., (2008) Simultaneous Fragmentation of Multiple Ions Using IMS Drift Time Dependent Collision Energies. *J. Am. Soc. Mass Spectrom.*, 19, 411-419.

Burgess, E.F. et al. (2008) Prostate cancer serum biomarker discovery through proteomic analysis of alpha-2 macroglobulin protein complexes. *Proteomics- Clinical Applications* 2, 1223-1233.

Cao X., Nesvizhskii A.I. (2008) Improved sequence tag generation method for peptide identification in tandem mass spectrometry. *J Proteome Res.*, Oct;7 (10):4422-34.



Cao Z., Li C., Higginbotham J.N., Franklin J.L., Tabb D.L., Graves-Deal R., Hill S., Cheek K., Jerome W.G., Lapierre L.A., Goldenring J.R., Ham A.J., Coffey R.J. (2008) Use of fluorescence-activated vesicle sorting for isolation of Naked2-associated, basolaterally targeted exocytic vesicles for proteomics analysis. *Mol Cell Proteomics*, Sep;7 (9):1651-67.

Carr, Steven A. and Anderson, Leigh (2008) Protein Quantitation through Targeted Mass Spectrometry: The Way Out of Biomarker Purgatory? *Clinical Chemistry*, 54:11, 1749–1752.

Cazares, L.H., Troyer, D., Mendaros, S., Lance, R.S., Nyalwidhe, J.O., Beydoun, H.A., Clements, M.A., Drake, R.R. and Semmes, O.J. (2009) Imaging Mass Spectrometry of a Specific Fragment of MEKK2 Discriminates Cancer from Uninvolved Prostate Tissue. *Clin. Cancer Res.*, *in press*.

Cho W., Jung K., Regnier F.E. (2008) Use of Glycan Targeting Antibodies to Identify Cancer-Associated Glycoproteins in Plasma of Breast Cancer Patients. *Analytical Chemistry*, 80(14), 5286-5292.

Choi H., Fermin D., Nesvizhskii A.I., (2008) Significance analysis of spectral count data in label-free shotgun proteomics. *Mol. Cell. Proteomics*, 7, 2373-2385.

Choi H., Ghosh D., Nesvizhskii A.I. (2008) Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling. *J Proteome Res.*, Jan;7 (1):286-92.

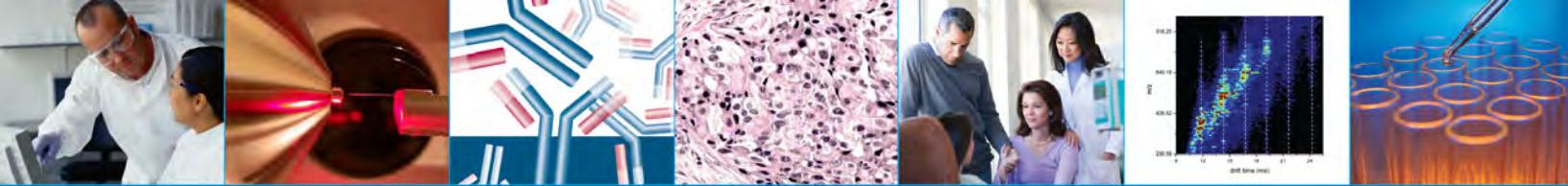
Choi H., Nesvizhskii A.I. (2008) False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J Proteome Res.*, Jan;7 (1):47-50.

Choi H., Nesvizhskii A.I. (2008) Semi-supervised model-based validation of peptide identifications in mass spectrometry-based proteomics. *J Proteome Res.*, Jan;7 (1):254-65.

Clowers B.H., Belov M.E., Prior D.C., Danielson W.F. 3rd, Ibrahim Y., Smith R.D. (2008) Pseudorandom sequence modifications for ion mobility orthogonal time-of-flight mass spectrometry. *Anal Chem.*, Apr 1;80 (7):2464-73.

Clowers B.H., Ibrahim Y.M., Prior D.C., Danielson W.F. 3rd, Belov M.E., Smith R.D. (2008) Enhanced ion utilization efficiency using an electrodynamic ion funnel trap as an injection mechanism for ion mobility spectrometry. *Anal Chem.*, Feb 1;80 (3):612-23.

Dalkic E., Nash D.E.W., Fassia M.K., and Chan C. (2009) Integrative Analysis of Cancer Pathway Progression. *Proteomics – Clinical Applications*, 3:473-85.



Ding Y., H. C Choi, and A.I. Nesvizhskii, (2008) Adaptive discriminant function analysis and re-ranking of MS/MS database search results for improved peptide identification in shotgun proteomics. *J. Proteome Res.*, 7, 4878-79.

Fenyo D., Phinney B.S., Beavis R.C. (2007) Determining the overall merit of protein identification data sets: rho-diagrams and rho-scores. *J Proteome Res.*, 6:1997-2004.

Edwards N.J. (2007) Novel Peptide Identification using ESTs and Sequence Database Compression. *Molecular Systems Biology*, 3.102.

Edwards N., Wu X., and Tseng C.-W. (2009) An Unsupervised, Model-Free, Machine-Learning Combiner for Peptide Identifications from Tandem Mass Spectra. *Clinical Proteomics*, 5.1.

Eriksson J., Fenyo D. (2007) Improving the success rate of proteome analysis by modeling protein-abundance distributions and experimental designs. *Nat Biotechnol.*, 25:651-5.

Fedulova I., Ouyang Z., Buck C. and Zhang X. (2007) PepTiger: Search Engine for Error-Tolerant Protein Identification from de Novo Sequences. *The Open Spectroscopy Journal*, 1, 1-8.

Feng J., Wong K.Y., Lynch G.C., Gao X., Pettitt B.M. (2007) Peptide conformations for a microarray surface-tethered epitope of the tumor suppressor p53. *J Phys Chem B*, Dec 13;111 (49):13797-806.

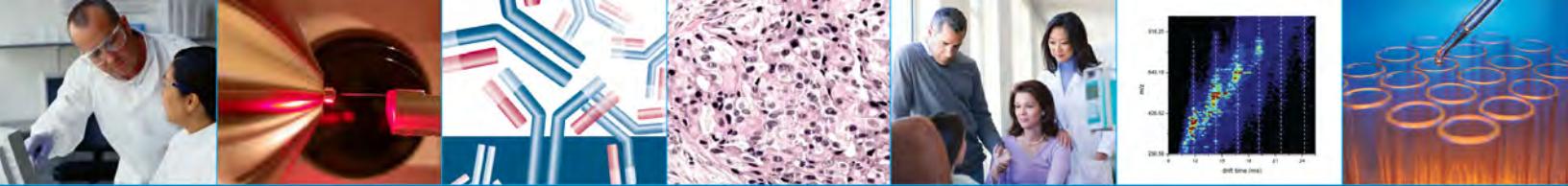
Fenselau C., Havey C., Teerakulkittipong N., Swatkoski S., Laine O., Edwards N. (2008) Identification of beta-lactamase in antibiotic-resistant *Bacillus cereus* spores. *Appl Environ Microbiol.*, Feb;74 (3):904-6.

Fenyo D., Phinney B.S., Beavis R.C. (2007) Determining the overall merit of protein identification data sets: rho-diagrams and rho-scores. *J Proteome Res.*, May;6 (5):1997-2004.

Fenyo D., Beavis R.C. (2008) Informatics development: challenges and solutions for MALDI mass spectrometry. *Mass Spectrom Rev.*, 27:1-19.

Fusaro, Vincent A.; Mani D.R.; Mesirov, Jill P.; and Carr, Steven A. (2009) Computational Prediction of High Responding Peptides for Development of Targeted Protein Assays. *Mass Spectrometry Nature Biotechnology*, *in press*.

Gatlin-Bunai C.L., Cazares L.H., Cooke W.E., Semmes O.J., Malyarenko D.I. (2007) Optimization of MALDI-TOF MS detection for enhanced sensitivity of affinity-captured proteins spanning a 100 kDa mass range. *J Proteome Res.*, Nov;6 (11):4517-24.



Gehrke, A.S., Sun, S., Kurgan, L., Ahn, N., Resing, K., Kafadar, K., and Cios, K. (2008) Improved machine learning method for analysis of gas phase chemistry of peptides. *BMC Bioinformatics*, 9:515-529.

Gong W., Zhou D., Ren Y., Wang Y., Zuo Z., Shen Y., Xiao F., Zhu Q., Hong A., Zhou X., Gao X., Li T. (2008) PepCyber:P~PEP: a database of human protein protein interactions mediated by phosphoprotein-binding domains. *Nucleic Acids Res.*, Jan;36 (Database issue):D679-83.

Goudreault M., L.M. D'Ambrosio, M. Mullin, B.G. Larsen, A. Sanchez, S. Chaudhry, G.I. Chen, F. Sicheri, A.I. Nesvizhskii, R. Aebersold, B. Raught, and A.C. Gingras, (2009) A PP2A phosphatase high-density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein. *Mol. Cell. Proteomics*, 8, 157-71.

Gredell J.A., Kini H.K. and Walton S.P. (2009) Engineering Cell Function by RNA Interference. *Cell Engineering*, Volume 6: Cell Line Development and Engineering.

Gredell J.A., Berger A.K., Walton S.P. (2008) Impact of target mRNA structure on siRNA silencing efficiency: A large-scale study. *Biotechnol Bioeng.*, Jul 1;100 (4):744-55.

Grote T., Siwak D.R., Fritzsche H., Joy C., Mills G.B., Simeone D., Whitcomb D.C., and Logsdon C.D. (2008) Reverse phase protein array for biomarkers in serum and plasma: Improved detection of CA19-9 in pancreatic cancer patient samples. *Proteomics*, 8:3051-3060.

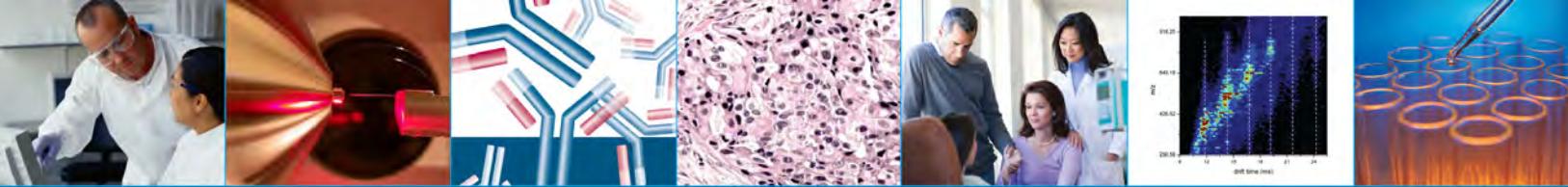
Grote T., Siwak D.R., Fritzsche H.A., Joy C., Mills G.B., Simeone D., Whitcomb D.C., Logsdon C.D. (2008) Validation of reverse phase protein array for practical screening of potential biomarkers in serum and plasma: accurate detection of CA19-9 levels in pancreatic cancer. *Proteomics*, Aug;8 (15):3051-60.

Ho L., J. Rohan, J. Wu, B. Staahl ,L. Chen, A. Kuo, J. Lessard, A. I. Nesvizhskii, J. Ranish, G. Crabtree (2009) Embryonic Stem Cell Chromatin Remodeling Complex, esBAF is Essential for Embryonic Stem Cell Self-Renewal and Pluripotency. *Proc. Nat. Acad. Sciences U.S.A.*, 106, 5181-6.

Hortin G.L., Sviridov D., Anderson N.L. (2008) High-Abundance Polypeptides of the Human Plasma Proteome Comprising the Top 4 Logs of Polypeptide Abundance. *Clinical Chemistry*, 54: 1608-1616.

Hu J., He X., Baggerly K.A., Coombes K.R., Hennessy B.T., Mills G.B. (2007) Non-parametric quantification of protein lysate arrays. *Bioinformatics*, Aug 1;23 (15):1986-94.

Ibrahim Y., Belov M.E., Tolmachev A.V., Prior D.C., Smith R.D. (2007) Ion funnel trap interface for orthogonal time-of-flight mass spectrometry. *Anal Chem.*, Oct 15;79 (20):7845-52.



Jaffe J.D., Keshishian H., Chang B., Addona T.A., Gillette M.A., Carr S.A. (2008) Accurate inclusion mass screening: a bridge from unbiased discovery to targeted assay development for biomarker verification. *Mol Cell Proteomics*, Oct; 7 (10):1952-62.

Jaffe J.D., Keshishian H., Chang B., Addona T.A., Gillette M.A., Carr S.A. (2008) Accurate Inclusion Mass Screening: a Bridge from Unbiased Discovery to Targeted Assay Development for Biomarker Verification. *Mol Cell Proteomics*, 7: 1952–1962.

Jung, K., Cho, W., Regnier, F.E. (2009) Glycoproteomics of Plasms Based On Narrow Selectivity Lectins. *J. Proteome Res.*, Accepted.

Karbassi, I.D., Cazares, L.H, Lance, R.S., Semmes, O.J and Drake, R.R. (2009) Proteomic Expression Profiling and Identification of Serum Proteins Using Immobilized Trypsin Beads with MALDI-TOF/TOF. *J. Proteome Res.*, in press.

Kelly R.T., Page J.S., Tang K., Smith R.D. (2007) Array of chemically etched fused-silica emitters for improving the sensitivity and quantitation of electrospray ionization mass spectrometry. *Anal Chem.*, Jun 1;79 (11):4192-8. Epub May 2.

Kelly R.T., Page J.S., Zhao R., Qian W.J., Mottaz H.M., Tang K., Smith R.D. (2008) Capillary-based multi nanoelectrospray emitters: improvements in ion transmission efficiency and implementation with capillary reversed-phase LC-ESI-MS. *Anal Chem.*, Jan 1;80 (1):143-9.

Keshishian H., Addona T., Burgess M., Kuhn E., Carr S.A. (2007) Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics*, Dec;6 (12):2212-29.

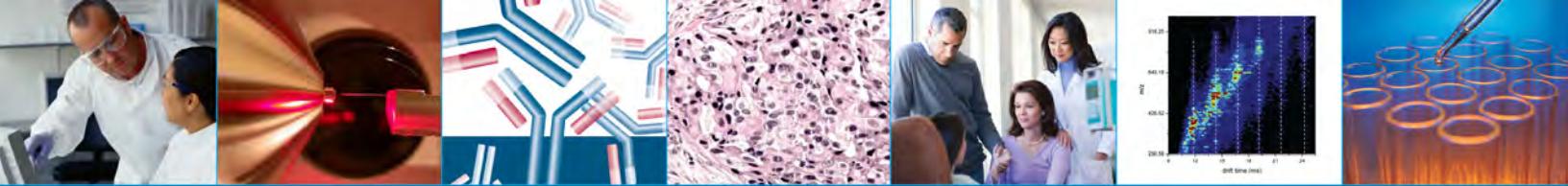
Kessner D., Chambers M., Burke R., Agus D., Mallick P. (2008) ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics*, 24:2534-2536.

Kim, B., Nesvizhskii, A.I., Reedy, P.G., Hahn, S., Aebersold, R., Ranish, J.A., (2007) The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Nat. Acad. Sciences U.S.A.*, 104, 16068-16073.

Kini H.K., Walton S.P. (2007) In vitro binding of single-stranded RNA by human Dicer. *FEBS Lett.*, Dec 11;581 (29):5611-6.

Kuhn, Eric; Addona, Terri; Keshishian, Hasmik; Burgess, Michael; Mani D.R., Sabatine, Marc S.; Gerszten, Robert E.; Carr, Steven A.. (2009) Developing Multiplexed Assays for Troponin I and Interleukin 33 in Plasma by Peptide Immunoaffinity Enrichment and Targeted Mass Spectrometry. *Clinical Chemistry*.

Kumar K.G., Barriere H., Carbone C.J., Liu J., Swaminathan G., Xu P., Li Y., Baker D.P., Peng J., Lukacs G.L., Fuchs S.Y. (2007) Site-specific ubiquitination exposes a linear motif to promote interferon-{alpha} receptor endocytosis. *J Cell Biol*, 179:935-50.



Kurulugama R.T., Valentine S.J., Sowell R.A., Clemmer D.E. (2008) Development of a high-throughput IMS-IMS-MS approach for analyzing mixtures of biomolecules. *J Proteomics*, Aug 21;71 (3):318-31.

Kuzyk, Michael A.; Smith, Derek; Yang, Juncong; Cross, Tyra; Jackson, Angela; Hardie, Darryl; Anderson, N. Leigh; Borchers, Christoph H. (2009) MRM-based, Multiplexed, Absolute Quantitation of 45 Proteins in Human Plasma. *Molecular and Cellular Proteomics*, Published online May 1. PMID: 19411661

Kuschner, Karl W. (2009) A Bayesian Network Approach to Feature Selection in Mass Spectrometry Data. *Doctoral Dissertation, College of William and Mary*, April, pp. 158, *in press*.

Lee, Bennett T.K.; Liew, Lailing; Lim, Jiahao; Tan, Jonathan K. L.; Lee, Tze Chuen; Veladandi, Pardha S.; Lim, Yun Ping; Han, Hao; Rajagopal, Gunaretnam; Anderson, N. Leigh (2008) Candidate List of yoUr Biomarker (CLUB): a Web-based Platform to Aid Cancer Biomarker Research. *Biomarker Insights*, 3:65–71.

Li Y.F., Arnold R.J., Li Y., Radivojac P., Sheng Q., Tang H. (2008) A Bayesian approach to protein inference problem in shotgun proteomics. The 12th Annual International Conference on Research in Computational Molecular Biology. *RECOMB*, pp. 167-180, Singapore.

Li Z., Srivastava S., Findlan R., Chan C. (2008) Using dynamic gene module map analysis to identify targets that modulate free fatty acid induced cytotoxicity. *Biotechnol Prog*, Jan-Feb;24 (1):29-37.

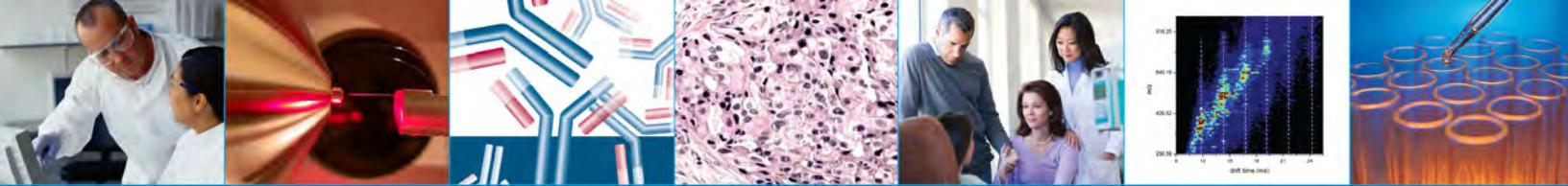
Lilja H., Ulmert D., Vickers A.J. (2008) Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat Rev Cancer*, 8:268-78.

Lilja H., Ulmert D., Vickers A.J. (2008) Systematic review of statistical methods used in molecular marker studies in cancer. *Cancer*, 112:1862-8.

Lilja H., Vickers A., Scardino P. (2007) Measurements of proteases or protease system components in blood to enhance prediction of disease risk or outcome in possible cancer. *J Clin Oncol*, 25:347-8.

Lin L., Williams B.A.R., Lindsay S.M. and Chaput J.C. Recognition Imaging Using a Synthetic Antibody to Gal80. *Chem.Phys.Chem.*, (manuscript in preparation)

Liu J., Kang S., Tang C., Ellis L.B., Li T. (2007) Meta-prediction of protein subcellular localization with reduced voting. *Nucleic Acids Res.*, 35 (15):e96.



Liu Z., Oh S.M., Okada M., Liu X., Cheng D., Peng J., Brat D.J., Sun S.Y., Zhou W., Gu W., Ye K. (2009) Human BRE1 is an E3 ubiquitin ligase for Ebp1 tumor suppressor. *Mol Biol Cell*, 20:757-68.

Lu Y., Adegoke O.A., Nepveu A., Nakayama K.I., Bedard N., Cheng D., Peng J., Wing S.S. (2009) USP19 deubiquitinating enzyme supports cell proliferation 1 by stabilizing KPC1, a ubiquitin ligase for p27Kip1. *Mol Cell Biolm.*, 29:547-58.

Marelli M., A.I. Nesvizhskii, and J.D. Aitchison, (2008) Identifying bona fide components of an organelle by isotope-coded labeling of subcellular fractions: an example in peroxisomes. *Methods Mol Biol.*, 432, 357-371.

Martin D.B., Holzman T., May D., Peterson A., Eastham A., Eng J., McIntosh M. (2008) MRMer, an interactive open source and cross-platform system for data extraction and visualization of multiple reaction monitoring experiments. *Mol Cell Proteomics*, Nov;7 (11):2270-8.

McConnell R.E., Higginbotham J.N., Shifrin D.A. Jr., Tabb D.L., Coffey R.J., Tyska M.J. (2009) The enterocyte microvillus is a vesicle-generating organelle. *J. Cell Biol.*, *in press*.

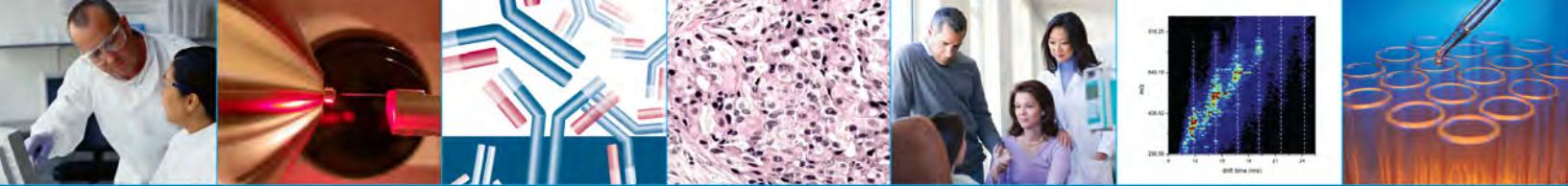
McLerran D.F., Feng Z., Semmes O.J., Cazares L., Randolph T.W. (2008) Signal detection in high-resolution mass spectrometry data. *J Proteome Res.*, Jan;7 (1):276-85.

McLerran D.F., W.E. Grizzle, Z. Feng, I.M. Thompson, W.L. Bigbee, W.H. Cazares, D.W. Chan, J. Dahlgren, J. Diaz, J. Kagan, D.W. Lin, G. Malik, D. Oelschlager, A. Partin, T.W. Randolph, L. Sokoll, S. Srivastava, S. Srivastava, M. Thornquist, D. Troyer, G.L. (2008) SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate Cancer. *Clinical Chemistry*, 54(1) 53-60.

McLerran D.F., W.E. Grizzle, Z. Feng, W.L. Bigbee, L.L. Banez, L.H. Cazares, D.W. Chan, J. Diaz, E. Izbicka, J. Kagan, D.E. Malehorn, G. Malik, D. Oelschlager, A. Partin, T.W. Randolph, N. Rosenzweig; S. Srivastava, S. Srivastava, I.M. Thompson, M.D. Thor (2008) Analytical Validation of Serum Proteomic Profiling for Diagnosis of Prostate Cancer: Sources of Sample Bias. *Clinical Chemistry*, 54(1) 44-52.

Mehrotra S., Lee I. and Chan C. (2009) pH Responsive Multilayer mediated Top-Down Forward Patterned siRNA Transfection with Linear Polyethylenimine-siRNA, Extended N/P Ratio Evaluation for Efficient Delivery. *Acta Biomaterialia*, 5:1474–1488.

Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J.L., and Slany, R.K., (2007) A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood*, 110, 4445-4454.



Nesvizhskii, A.I. ( 2007) Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol Biol.*, 367, 87-119.

Nesvizhskii, A.I., Vitek, O., Aebersold, R. (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods*, 4, 787-797.

Oh C. , ZZak S., Mirzaei H., Buck C., Regnier F., Zhang X. (2007) Neural network prediction of peptide separation in strong anion exchange chromatography. *Bioinformatics*, 23 (1):114-118.

Old, W.M., Shabb, J.B., Houel, S., Wang, H., Couts, K.L., Yen, C.-y., Litman, E. S., Croy, C.H., Meyer-Arendt, K., Miranda, J.G., Brown, R.A., Witze, E.S., Schweppe, R.E., Resing, K.A., and Ahn, N.G. (2009) Functional Proteomics Identifies Targets of Phosphorylation by B-Raf Signaling in Melanoma. *Molecular Cell*, 34: 115-131.

Ott, L.W., Resing, K.A., Sizemore, A.W., Heyen, J.W., Cocklin, R.R., Pedrick, N.M., Woods, H.C., Chen, J.Y., Goebl, M.G., Witzmann, F.A., and Harrington, M.A. (2007) Tumor Necrosis Factor-alpha- and interleukin-1-induced cellular responses: coupling proteomic and genomic information. *J. Proteome Res.*, 6(6):2176-2185.

Padliya N.D., Garrett W.M., Campbell K.B., Tabb D.L., Cooper B. (2007) Tandem mass spectrometry for the detection of plant pathogenic fungi and the effects of database composition on protein inferences. *Proteomics*, Nov;7 (21):3932-42.

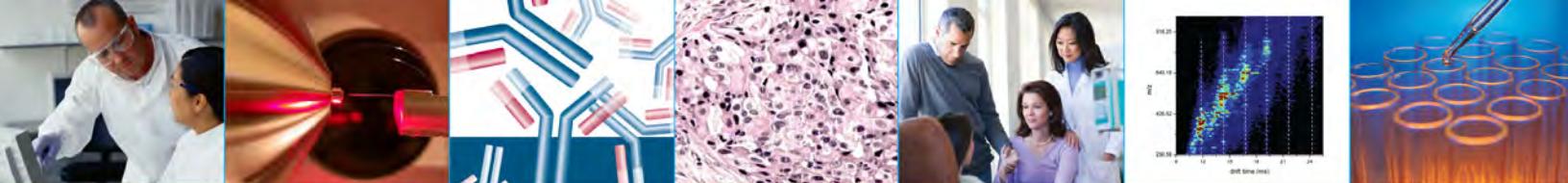
Page J.S., Kelly R.T., Tang K., Smith R.D. (2007) Ionization and transmission efficiency in an electrospray ionization-mass spectrometry interface. *J Am Soc Mass Spectrom.*, Sep;18 (9):1582-90.

Page J.S., Marginean I., Baker E.S., Kelly R.T., Tang K., Smith R.D. (2008) Biases in ion transmission through an electrospray ionization – mass spectrometry capillary inlet. Submitted to *J. Am. Soc. Mass Spectrom.*

Paulovich A.G., D. Billheimer, A.-J.L. Ham, L.J. Vega-Montoto, P.A. Rudnick, D.L. Tabb, P. Wang, R.K. Blackman, D.M. Bunk, H.L. Cardasis, K.R. Clouser, C.R. Kinsinger, B. Schilling, T.J. Tegeler, A.M. Variyath, M. Wang, J.R. Whiteaker, L.J. Zimmerman, S.A. Carr, S.J. Fisher, B.W. Gibson, M. Mesri, T.A. Neubert, F.E. Regnier, H. Rodriguez, C. Spiegelman, S.E. Stein, P. Tempst, D.C. Liebler, and the NCI Clinical Proteomic Technology Assessment for Cancer (CPTAC) Network. A Reference Yeast Proteome for Benchmarking LC-MS Platform Performance. submitted to *Nature Biotechnology*.

Paulovich A.G., Whiteaker J.R., Hoofnagle A.N., and Wang P. (2008) The Interface between Biomarker Discovery and Clinical Validation: The Tar Pit of the Protein Biomarker Pipeline. *Proteomics-Clinical Applications*, 2: 1386–1402.

Peng J. (2008) Evaluation of proteomic strategies for analyzing ubiquitinated proteins. *BMB Rep.*, Mar 31;41 (3):177-83.



Peng I.X., Shiea J., Ogorzalek Loo R.R., Loo J.A. (2007) Electrospray-assisted laser desorption/ionization and tandem mass spectrometry of peptides and proteins. *Rapid Commun Mass Spectrom*, 21 (16):2541-6.

Peng I.X., Ogorzalek Loo R.R., Shiea J., Loo J.A. (2008) Reactive-electrospray-assisted laser desorption/ionization for characterization of peptides and proteins. *Anal Chem.*, Sep 15;80 (18):6995-7003.

Pope, Matthew E.; Soste, Martin V.; Eyford Brett A.; Anderson, N. Leigh and Pearson, Terry W. (2009) Anti-peptide antibody screening: selection of high affinity monoclonal reagents by a refined surface plasmon resonance technique. *J Immunol Methods*, 341: 86-96.

Qiu J., G. Choi, L. Li, H. Wang, S.J. Pitteri, S.R. Pereira-Faca, A.L. Krasnoselsky, T.W. Randolph, C. Edelstein, M. Barnett, M. Thornquist, G. Goodman, G. Omenn, D. Brenner, Z. Feng and S.M. Hanash (2008) Occurrence of autoantibodies to annexin 1, 14-3-3 theta and LAMR1 in pre-diagnostic lung cancer sera. *Journal of Clinical Oncology*, Nov. 1, 5060—5066.

Qiu R. and Regnier F. (2005) Use of Multidimensional Lectin Affinity Chromatography in Differential Glycoproteomics. *Anal. Chem.*, 77, 2802-2809.

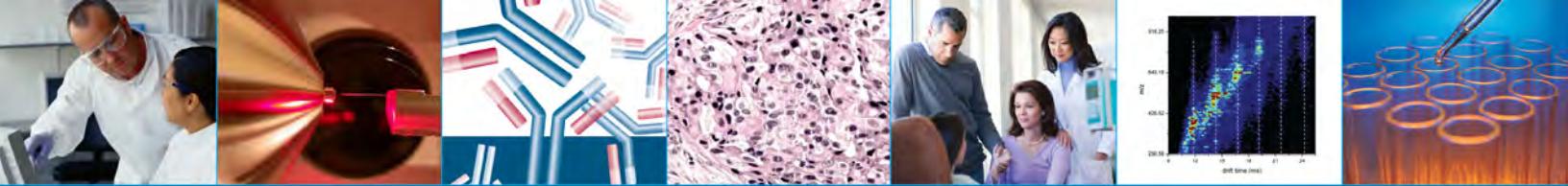
Qiu R. and Regnier F. (2005) Comparative Glycoproteomics of N-Linked Complex-Type Glycoforms Containing Sialic Acid in Human Serum. *Anal. Chem.*, 77, 7225-7231.

Qiu R., Zhang X., Regnier F. (2007) A method for the identification of glycoproteins from human serum by a combination of lectin affinity chromatography along with anion exchange and Cu-IMAC selection of tryptic peptides. *Journal of Chromatography B.*, 845, 143–150.

Ramesh Babu J., Lamar Seibenhener M., Peng J., Strom A.L., Kemppainen R., Cox N., Zhu H., Wooten M.C., Diaz-Meco M.T., Moscat J., Wooten M.W. (2008) Genetic inactivation of p62 leads to accumulation of hyperphosphorylated tau and neurodegeneration. *J Neurochem.*, 106:107-20.

Raphael B.J., Volik S., Yu P., Wu C., Huang G., Linardopoulou E.V., Trask B.J., Waldman F., Costello J., Pienta K.J., Mills G.B., Bajsarowicz K., Kobayashi Y., Sridharan S., Paris P.L., Tao Q., Aerni S.J., Brown R.P., Bashir A., Gray J.W., Cheng J.F., de Jong P., Nefedov M., Ried T., Padilla-Nash H.M., Collins C.C. (2008) A sequence-based survey of the complex structural organization of tumor genomes. *Genome Biol.*, 9 (3):R59.

Riley C. P., Gough, E. S., He, J., Jandhyala, S., Kennedy, B., Orcun, S., Ouzzani, M., Buck, C. R., Zhang, X. (2009) The proteome discovery pipeline – a data analysis pipeline for mass spectrometry-based differential proteomics. *BMC Bioinformatics*, Accepted.



Robinson S., Niles R.K., Witkowska H.E., Rittenbach K.J., Nichols R.J., Sargent J.A., Dixon S.E., Prakobphol A., Hall S.C., Fisher S.J., Hardt M. (2008) A mass spectrometry-based strategy for detecting and characterizing endogenous proteinase activities in complex biological samples. *Proteomics*, Feb;8 (3):435-45.

Rodriguez H., Snyder M., Uhlén M., Andrews P., Beavis R.C., Borchers C.H., Chalkley R., Cho S.Y., Cottingham K., Dunn M., Dylag T., Edgar R., Hare P., Heck A.J., Hirsch R.F., Kennedy K., Kolar P., Kraus H.J., Mallick P., Nesvizhskii A., Ping P., Ponten F., Yang L., Yates J.R., Stein S.E., Hermjakob H., Kinsinger C.R., Apweiler R. (2009) *J. Proteome Res.*, *in press*.

Rudnick P.A., K.R. Clauser, L.E. Kilpatrick, D.V. Tchekhovskoi, P. Neta, N. Blonder, D.D. Billheimer, R.K. Blackman, D.M. Bunk, H.L. Cardasis, A.-J.L. Ham, J.D. Jaffe, C.R. Kinsinger, M. Mesri, T.A. Neubert, B. Schilling, D.L. Tabb, T.J. Tegeler, L. Vega-Montoto, A.M. Variyath, M. Wang, P. Wang, J.R. Whiteaker, L.J. Zimmerman, S.A. Carr, S.J. Fisher, B.W. Gibson, T.A. Neubert, A.G. Paulovich, F.E. Regnier, H. Rodriguez, C. Spiegelman, P. Tempst, D.C. Liebler, S.E. Stein and the CPTAC Network. Performance Metrics for Evaluating Liquid Chromatography-Tandem Mass Spectrometry Systems in Shotgun Proteomics. submitted to *Nature Methods*.

Rudomin, E., Carr, S.A., Jaffe, J. (2009) Directed Sample Interrogation Utilizing an Accurate Mass Exclusion-Based Date-Dependent Acquisition Strategy (AMEx). *Journal of Proteome Research*.

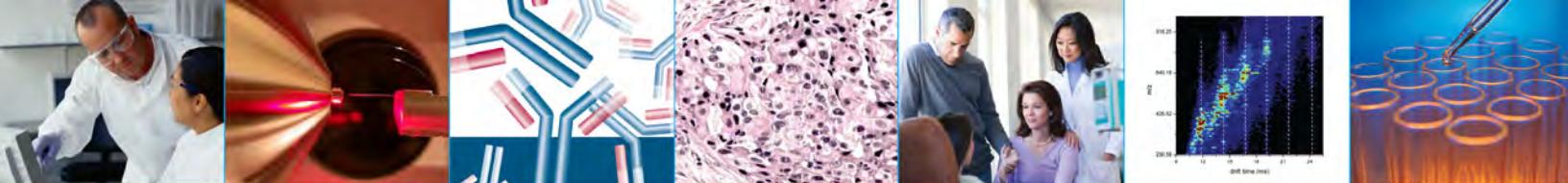
Saha, S., Harrison, S., Chen, J. (2008) Dissecting the Human Plasma Proteome and Inflammatory Response Biomarkers. PROTEOMICS. *Clinical Applications*, PRCA-2008-00044-R1.

Saha, S., Harrison, S., Shen, C., Tang, H., Radivojac, P., Arnold, R., Zhang, X., Chen, J.Y. (2008) HIP2: An online database of human plasma proteins from healthy individuals. *BMC Medical Genomics*, 1:12 doi:10.1186/1755-8794-1-12.

Searle B.C., Turner M., Nesvizhskii A.I. (2008) Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. *J Proteome Res.*, 7 (1):245-53.

Seyfried N.T., Xu P., Duong D.M., Cheng D., Hanfelt J., Peng J. (2008) Systematic approach for validating the ubiquitinated proteome. *Anal Chem.*, Jun 1;80 (11):4161-9.

Shen C., Wang Z., Shankar G., Zhang X., Li L. (2008) A hierarchical statistical model to assess the confidence of peptides and proteins inferred from tandem mass spectrometry. *Bioinformatics*, Jan 15;24 (2):202-8.



Shen C., Wang Z., Shankar G., Zhang X., and Li L. (2007) A Hierarchical Statistical Model to Assess the Confidence of Peptides and Proteins Inferred from Tandem Mass Spectrometry. *Bioinformatics Advance Access, published November 17.*

Slebos R.J., Brock J.W., Winters N.F., Stuart S.R., Martinez M.A., Li M., Chambers M.C., Zimmerman L.J., Ham A.J., Tabb D.L., and Liebler D.C. (2008) Evaluation of Strong Cation Exchange versus Isoelectric Focusing of Peptides for Multidimensional Liquid Chromatography-Tandem Mass Spectrometry. *J Proteome Res.*

Sours K.M., Kwok S.C., Rachidi T., Lee T., Ring A., Hoofnagle A.N., Resing K.A., Ahn N.G. (2008) Hydrogen-exchange mass spectrometry reveals activation-induced changes in the conformational mobility of p38alpha MAP kinase *J Mol Biol.*, Jun 20;379 (5):1075-93.

Sprung R.W., Brock J.W.C., Tanksley J.P., Li M., Washington M.K., Slebos R.J.C., Liebler D.C. (2009) Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol. Cell. Proteomics* in press.

Srivastava S., Li Z., Yang X., Yedwabnick M., Shaw S., Chan C. (2007) Identification of genes that regulate multiple cellular processes/responses in the context of lipotoxicity to hepatoma cells. *BMC Genomics*, Oct 9;8:364.

Srivastava S., Zhang L., Jin R., Chan C. (2008) A novel method incorporating gene ontology information for unsupervised clustering and feature selection. *PLoS ONE*, 3 (12):e3860.

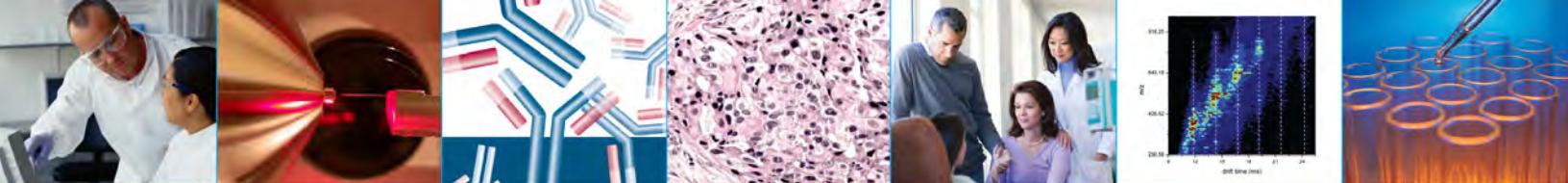
Sun, S., Meyer-Arendt, K., Eichelberger, B., Brown, R., Yen, C.Y., Old, W.M., Pierce, K., Cios, K., Ahn, N.G., and Resing, K.A. (2007) Improved validation of peptide MS/MS assignments using spectral intensity prediction and full annotation of fragment ions. *Mol. Cell. Proteomics*, 6:1-17.

Swatkoski S., Gutierrez P., Ginter J., Petrov A., Dinman J.D., Edwards N., and Fenselau C. (2007) Integration of residue specific acid cleavage into proteomic workflows. *Journal of Proteome Research*, 6.11:4525-4527.

Swatkoski S., Gutierrez P., Wynne C., Petrov A., Dinman J.D., Edwards N., and Fenselau C. (2008) Evaluation of microwave accelerated residue-specific acid cleavage for proteomic applications. *Journal of Proteome Research*, 7.2:579-586.

Tabb D.L. (2008) What's driving false discovery rates?. *J Proteome Res.*, Jan;7 (1):45-6.

Tabb D.L., Fernando C.G., Chambers M.C. (2007) MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res.*, Feb;6 (2):654-61.



Tabb D.L., Ma Z.Q., Martin D.B., Ham A.J., Chambers M.C. (2008) DirecTag: accurate sequence tags from peptide MS/MS through statistical scoring. *J Proteome Res.*, Sep;7 (9):3838-46

Taylor B.S., M. Pal, J. Yu, B. Laxman, S. Sundaram, R. Zhao, A. Menon, J.T. Wei, A.I. Nesvizhskii, D. Ghosh, G.S. Omenn, D.M. Lubman, A.M. Chinnaiyan, and A. Sreekumar, (2008) Humoral response profiling reveals pathways to prostate cancer progression. *Mol. Cell. Proteomics*, 7, 600-611.

Taylor C.F., Paton N.W., Lilley K.S., Binz P.A., Julian R.K. Jr., Jones A.R., Zhu W., Apweiler R., Aebersold R., Deutsch E.W., Dunn M.J., Heck A.J., Leitner A., Macht M., Mann M., Martens L., Neubert T.A., Patterson S.D., Ping P., Seymour S.L., Souda P., Tsugita A., Vandekerckhove J., Vondriska T.M., Whitelegge J.P., Wilkins M.R., Xenarios I., Yates J.R. 3rd, Hermjakob H. (2007) The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol.*, 25:887-93.

Taylor C.F., Binz P.A., Aebersold R., Affolter M., Barkovich R., Deutsch E.W., Horn D.M., Hühmer A., Kussmann M., Lilley K., Macht M., Mann M., Müller D., Neubert T.A., Nickson J., Patterson S.D., Raso R., Resing K., Seymour S.L., Tsugita A., Xenarios I., Zeng R., Julian R.K. Jr. (2008) Guidelines for reporting the use of mass spectrometry in proteomics. *Nat Biotechnol.*, 26:860-1.

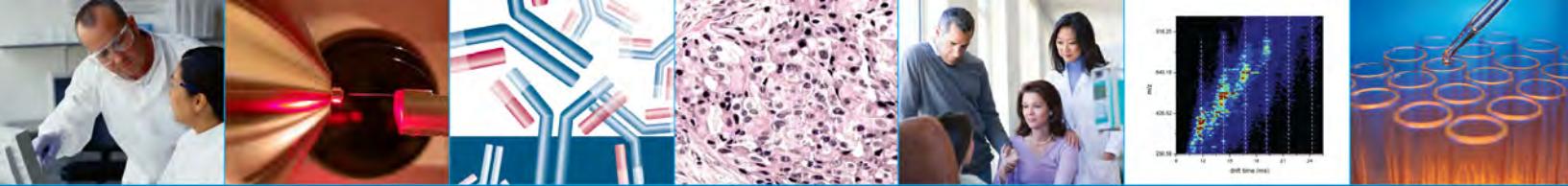
Tracy M.B., Chen H., Weaver D.M., Malyarenko D.I., Sasinowski M., Cazares L.H., Drake R.R., Semmes O.J., Tracy E.R., Cooke W.E. (2008) Precision enhancement of MALDI-TOF MS using high resolution peak detection and label-free alignment. *Proteomics*, Apr;8 (8):1530-8.

Ullintz P.J., B. Bodenmiller, R. Aebersold, P.C. Andrews, and A.I. Nesvizhskii, (2008) Investigating MS<sup>2</sup>-MS<sup>3</sup> matching statistics: A model for coupling consecutive stage mass spectrometry data for increased peptide identification confidence. *Mol. Cell. Proteomics*, 7, 71-87.

Ullintz P.J., Bodenmiller B., Andrews P.C., Aebersold R., Nesvizhskii A.I. (2008) Investigating MS2/MS3 matching statistics: a model for coupling consecutive stage mass spectrometry data for increased peptide identification confidence. *Mol Cell Proteomics*, Jan;7 (1):71-87.

Ullintz P.J., A. Yocum, B. Bodenmiller, R. Aebersold, P.C. Andrews, and A.I. Nesvizhskii (2009) Comparison of MS<sup>2</sup>-only, MSA, and MS<sup>2</sup>/MS<sup>3</sup> methodologies for phosphopeptide Identification. *J. Proteome Res.*, 8, 887-99.

Varghese B., Barriere H., Carbone C.J., Banerjee A., Swaminathan G., Plotnikov A., Xu P., Peng J., Goffin V., Lukacs G.G., Fuchs S.Y. (2008) Polyubiquitination of prolactin receptor stimulates its internalization, post-internalization sorting, and degradation via the lysosomal pathway. *Mol Cell Biol*, 28:5275-87.



Vickers A.J., Ulmert D., Serio A.M., Björk T., Scardino P.T., Eastham J.A., Berglund G., Lilja H. (2007) The predictive value of prostate cancer biomarkers depends on age and time to diagnosis: towards a biologically-based screening strategy. *Int J Cancer*, 121:2212-7.

Villanueva J., Lawlor K., Toledo-Crow R., Tempst P. (2006) Automated serum peptide profiling. *Nature Protocols*, 1:880-891.

Villanueva J., Martorella A.J., Lawlor K., Philip J., Fleisher M., Robbins R.J., Tempst P. (2006) Serum peptidome patterns that distinguish metastatic thyroid carcinoma from cancer-free controls are unbiased by gender and age. *Mol Cell Proteomics*, 5:1840-1852.

Villanueva J., Nazarian A., Lawlor K., Tempst P. (2009) Monitoring peptidase activities in complex proteomes by MALDI-TOF mass spectrometry. *Nature Protocols*, in press.

Villanueva J., Nazarian A., Lawlor K., Yi S.S., Robbins R.J., Tempst P. (2008) A sequence-specific exopeptidase activity test (SSEAT) for "functional" biomarker discovery. *Mol Cell Proteomics*, Mar;7 (3):509-18.

Villanueva J., Philip J., DeNoyer L., Tempst P. (2007) Data analysis of assorted serum peptidome profiles. *Nature Protocols*, 2:588-602.

Vitzthum Frank, Hoerth G. Patric, Schulz-Knappe Peter, Tammen Harald, Adamkiewicz Juergen and Anderson N. Leigh (2007) Metrological sharp shooting for plasma proteins and peptides: The need for reference materials for accurate measurements in clinical proteomics and in vitro diagnostics to generate reliable results. *Proteomics Clin. Appl.*, 1: 1016-1035.

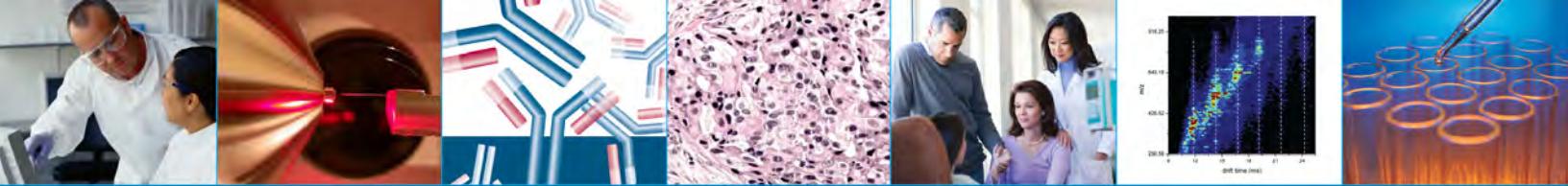
Walton S.P. and Jayaraman A. (2009) Proteomics: Technology Development and Applications. *Expert Review of Proteomics*, 6:23-25.

Wan J., Kang S., Tang C., Yan J., Ren Y., Liu J., Gao X., Banerjee A., Ellis L.B., Li T. (2008) Meta-prediction of phosphorylation sites with weighted voting and restricted grid search parameter selection. *Nucleic Acids Res.*, Mar;36 (4):e22.

Wang J., Gutierrez P., Edwards N., and Fenselau C. (2007) Integration of <sup>18</sup>O Labeling and Solution Isoelectric Focusing in a Shotgun Analysis of Mitochondrial Proteins. *Journal of Proteome Research*, 6.12: 4601-4607.

Wang P., Whiteaker J.R., and Paulovich A.G. (2009) The evolving role of mass spectrometry in cancer biomarker discovery. *Cancer Biology & Therapy*, In Press.

Wang X., Dalkic E., Wu M., and Chan C. (2008) Gene module level analysis: identification to networks and dynamics. *Curr Opin Biotechnol.*, 19:482-91.



Wang X., Dalkic E., Wu M., Chan C. (2008) Gene module level analysis: identification to networks and dynamics. *Current Opinion in Biotechnology*, October, 19 (5) Pages 482-491

Whiteaker J.R., Zhang H., Zhao L., Wang P., Kelly-Spratt K.S., Ivey R.G., Piening B.D., Feng L.C., Kasarda E., Gurley K.E., Eng J.K., Chodosh L.A., Kemp C.J., McIntosh M.W., Paulovich A.G. (2007) Integrated Pipeline for Mass Spectrometry-Based Discovery and Confirmation of Biomarkers Demonstrated in a Mouse Model of Breast Cancer. *Journal of Proteome Research*, 6(10):3962-75.

Whiteaker J.R., Zhao L., Zhang H.Y., Feng L., Piening B.D., Anderson L., and Paulovich A.G. (2007) Antibody-based enrichment of peptides on magnetic beads for mass spectrometry-based quantification of serum biomarkers. *Analytical Biochemistry*, 362(1):44-54.

Williams, B.A.R., Diehnelt, C.W., Belcher, P., Greving, M., Woodbury, N.W., Johnston, S.A., and Chaput, J.C. Synthetic Protein Affinity Reagents by Combining Peptide Ligands on Novel DNA Scaffolds. *Proc. Natl. Acad. Sci. USA*, (manuscript ready for submission).

Winborn B.J., Travis S.M., Todi S.V., Scaglione K.M., Xu P., Williams A.J., Cohen R.E., Peng J., Paulson H.L. (2008) The deubiquitinating enzyme ataxin-3, a polyglutamine disease protein, edits K63-linkages in mixed linkage ubiquitin chains. *J Biol Chem*, 283:26436-43.

Witze E.S., Old W.M., Resing K.A., Ahn N.G. (2007) Mapping protein post-translational modifications with mass spectrometry. *Nat Methods*, Oct;4 (10):798-806.

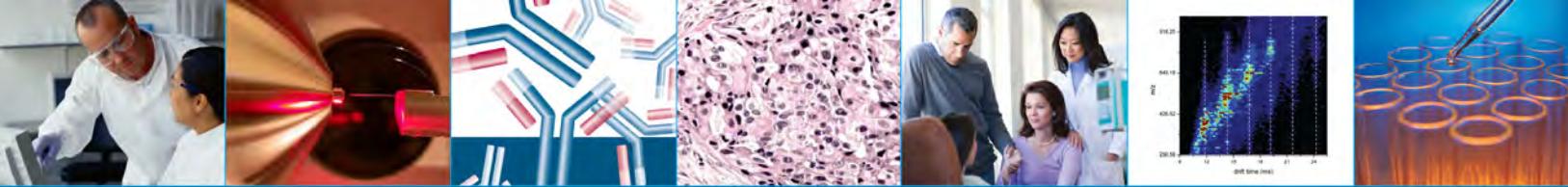
Wooten M.W., Geetha T., Babu J.R., Seibenhener M.L., Peng J., Cox N., Diaz-Meco M.T., Moscat J. (2008) Essential role of SQSTM1/p62 in regulating accumulation of K63-ubiquitinated proteins. *J Biol Chem*, 283:6783-9.

Wu X., Tseng C.-W., and Edwards N. (2007) HMMatch: Peptide Identification by Spectral Matching of Tandem Mass Spectra using Hidden Markov Models. *Journal of Computational Biology*, 14.8: 1025-1043.

Xie S. and Walton S.P. (2009) Application and analysis of structure-switching aptamers for small molecule quantification. *Analytica Chimica Acta*, 638:213-219.

Xiong L., D. Andrews, and F. Regnier (2003) Comparative Proteomics of Glycoproteins Based and Isotope Coding. *Journal of Proteome Research*, 2, 618-625.

Xu P., Duong D.M., Seyfried N.T., Cheng D., Xie Y., Robert J., Rush J., Hochstrasser M., Finley D., Peng J. (2009) Quantitative proteomics reveals the function of unconventional polyubiquitin chains in proteasomal degradation. *Cell*, 137:133-145.



Xu P., Peng J. (2008) Characterization of polyubiquitin chain structure by middle-down mass spectrometry. *Anal Chem.*, May 1;80 (9):3438-44.

Yang X. and Chan C. (2009) Repression of PKR Mediates Palmitate-Induced Apoptosis in HepG2 Cells through Bcl-2. *Cell Res.*, 19(4):469-486.

Yang X., Zhou Y., Jin R., and Chan C. (2009) Reconstruct modular phenotype-specific gene networks by knowledge-driven matrix factorization. *Bioinformatics*, (revision in review).

Yen, C. Y., Meyer-Arendt, K., Eichelberger, B., Sun, S., Houel, S., Old, W. M., Knight, R., Ahn, N. G., Hunter, L. E., and Resing, K. A. (2008) A simulated MS/MS library for spectrum-to-spectrum searching in large-scale identification of proteins. *Mol. Cellular Proteomics*, 8: 857-869.

Yoo, C., Patwa, T.H., Kreunin, P., Miller, F.R., Huber, G.C., Nesvizhskii, A.I., Lubman, D.M., (2007) Comprehensive analysis of proteins of pH fractionated samples using monolithic LC/MS/MS, intact MW measurement and MALDI-QIT-TOF MS. *J. Mass Spectrom.*, 42, 312-334.

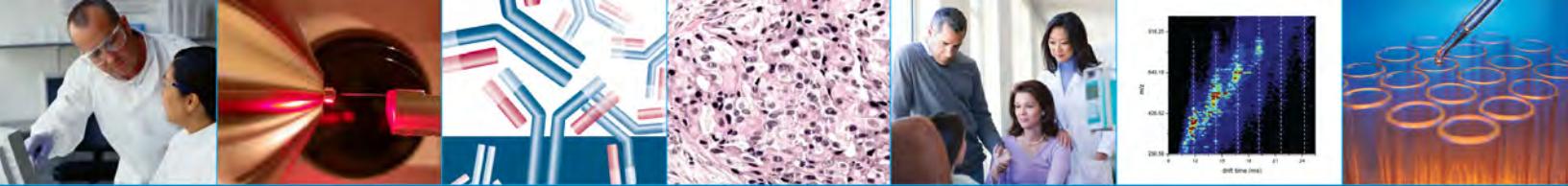
Zhang B., Chambers M.C., Tabb D.L. (2007) Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J Proteome Res.*, Sep;6 (9):3549-57. Epub Aug 4.

Zhang, J., Neilson, M., and Chaput, J.C. Peptides with High Affinity and Specificity Occur Frequently in Large Pools of Random Sequences. *ACS Chem. Biol.*, (manuscript in preparation).

Zhang, J., Williams, B., and Chaput, J.C. Synthesis of a Bivalent Affinity Reagent to Human Transferrin. *J. Am. Chem. Soc.*, (manuscript in preparation).

Zhang M., Ouyang Q., Stephenson A., Kane M., Salt D., Prabhakar S., Burgner J., Buck C. and Zhang X. (2008) Interactive analysis of systems biology molecular expression data. *BMC Systems Biology*, 2:23 doi:10.1186/1752-0509-2-23.

Zhang X., Oh C. Riley C.P., Buck, C. (2007) Current Status of Computational Approaches for Protein Identification Using Tandem Mass Spectra. *Current Proteomics*, 4, 121-130.



## 4.14 CPTC Patent Disclosures/Applications Listing

### **Chaput**

*Synthetic Antibodies*, AZTE Tech ID # M8-087L; PCT Authorized on 04/23/08.

### **Fisher**

The LBNL and UCSF groups filed a joint patent disclosure on the use of the MECA-79-reactive CD44 variant and other glycoproteins that carry this modification as biomarkers that enable the early detection of the basal subtype of breast cancer and as therapeutic targets for preventing metastases.

### **Gao**

Gao X., Zhou X., Zhang X., Hong A., Zue Q. "Make and use of surface molecules of varies densities" WO2008/003100 (filed in China and Europe)

### **Smith**

M.E. Belov, "Dynamic Multiplexed Methods using Ion Mobility Spectrometer" Filed 6/2008

M. E. Belov, B.H. Clowers, Y.M. Ibrahim, D.C. Prior, R.D. Smith, "Ion Funnel Ion Trap and Process" Filed 5/2008

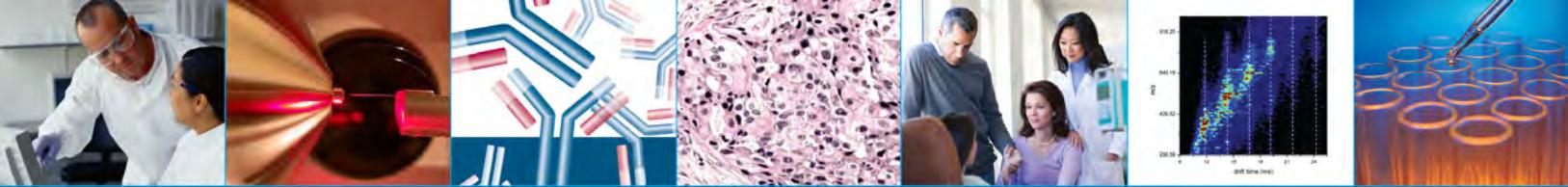
M.E. Belov, R.D. Smith, "Method of Multiplexed Analysis using Ion Mobility Spectrometer" filed 2/2007

### **Templeton**

D. Templeton, "Mass tagging for quantitative analysis of biomolecules using <sup>13</sup>C labeled phenylisocynate" Provisional Filed 2/15/06, revised 8/12/08.

### **Walton**

Xie, S. and Walton, S. P. "Proteomics Analyses using Labeled Aptamers." Disclosure and provisional filed.



## 4.15 Upcoming and Past Speaking Engagements for CPTC Leadership

CPTC is committed to engaging with members of the scientific community to advance the study of clinical cancer proteomics. Throughout the year, CPTC leadership speaks at national and international events about the important role that proteomic technologies will play in solving mission-critical problems in cancer research, including detecting cancer processes, finding targets for novel therapeutics, and determining biological markers of treatment response.

**2009**

### **October 7-9, 2009**

Innovative Molecular Analysis Technologies Annual Meeting  
Organized by: NCI Innovative Molecular Analysis Technologies  
Hyatt Regency Bethesda  
Bethesda, MD

### **October 5-7, 2009**

Clinical Proteomic Technologies for Cancer Annual Meeting  
Organized by: NCI Clinical Proteomic Technologies for Cancer  
Hyatt Regency Bethesda  
Bethesda, MD

### **September 22-24, 2009**

Translational Cancer Medicine: Optimizing Oncology Drug and Diagnostic Development Meeting  
Organized by: Cambridge Healthtech Institute  
Grand Hyatt Washington  
Washington, DC

### **June 1-4, 2009**

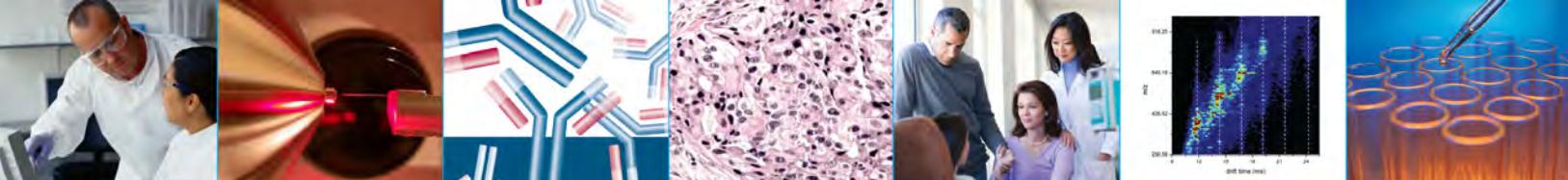
American Society for Mass Spectrometry Annual Meeting  
Organized by: American Society for Mass Spectrometry  
Philadelphia, PA

### **May 14, 2009**

Teleconference: Understanding NCI  
Organized by: NCI Office of Advocacy Relations  
Bethesda, MD

### **May 12-13, 2009**

International Workshop on Rapid, Pre-Publication Data Release



Organized by: Genome Canada  
Toronto, Canada

**April 28, 2009**

Translating Clinical Proteomics

Organized by: Center Department of Biochemistry & Molecular Biology, University of Louisville  
Louisville, KY

**April 18-22, 2009**

American Association for Cancer Research Annual Meeting

Organized by: American Association for Cancer Research  
Colorado Convention Center  
Denver, CO

**March 24-26, 2009**

Korean HUPO Annual Meeting

Organized by: Korean HUPO  
University Cultural Center, Seoul National University  
Seoul, Korea

**March 23-25, 2009**

Affinity Proteomics: 4th ESF Workshop on Ligand Binders against the Human Proteome

Organized by: European Science Foundation  
Congress Centre  
Alpbach, Austria

**March 24, 2009**

Teleconference: Steering Committee

Organized by: Foundation for the National Institutes of Health  
Bethesda, MD

**March 19, 2009**

Teleconference: Understanding NCI

Organized by: NCI Office of Advocacy Relations  
Bethesda, MD

**February 22-25, 2009**

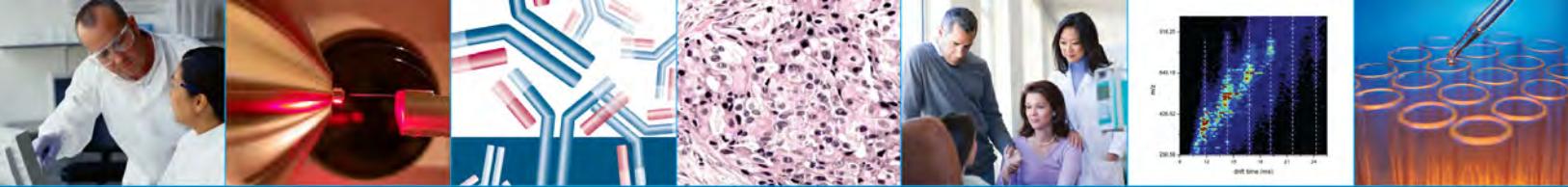
Integrative Proteomics for the Future

Organized by: US HUPO  
Westin Gaslamp District Hotel  
San Diego, CA

**February 19, 2009**

NHLBI Proteomics External Advisory Panel Meeting

Organized by: National Heart, Lung, and Blood Institute



Westin Gaslamp District Hotel  
San Diego, CA

**February 12, 2009**

Center for Environmental Genomics and Integrative Biology Symposia/Workshop  
Organized by: Center for Environmental Genomics and Integrative Biology University of Louisville  
Louisville, KY

**January 11-16, 2009**

Cambridge Healthtech Institute's 8th Annual PepTalk  
Organized by: Cambridge Healthtech Institute  
Hotel Del Coronado  
San Diego, CA

**2008**

**October 15-17, 2008**

4th Modern Drug Discovery and Development Summit (M3D)  
Organized by: GTCbio  
Hilton La Jolla Torrey Pines  
La Jolla, CA

**September 29-October 1, 2008**

Biomarker Discovery Summit  
Organized by: Cambridge Healthtech Institute  
Loews Philadelphia Hotel  
Philadelphia, PA

**August 16-20, 2008**

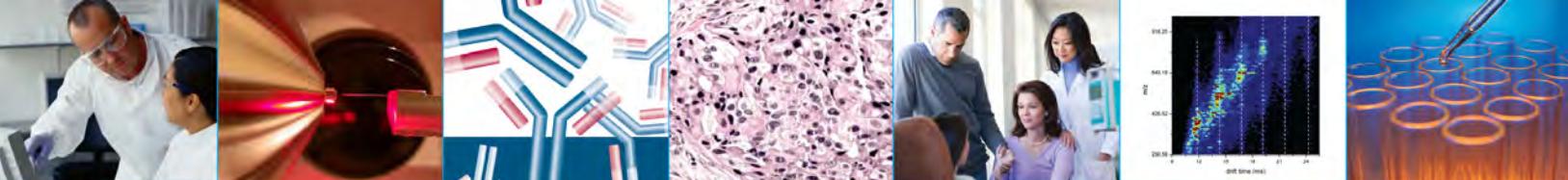
HUPO 2008 7th World Congress  
Organized by: Human Proteome Organization  
Amsterdam RAI Convention Centre  
Amsterdam, the Netherlands

**August 14, 2008**

International Summit on Proteomics Data Release and Sharing  
Hotel Novotel  
Amsterdam, the Netherlands

**August 4, 2008**

Proteomics Symposium



University of Southern California/Keck School of Medicine  
Aresty Auditorium, Harlyne Norris Cancer Tower  
Los Angeles, CA

**July 27-31, 2008**

American Association for Clinical Chemistry 2008 Annual Meeting  
Organized by: American Association for Clinical Chemistry  
Washington, DC

**July 10-11, 2008**

2nd Annual Rediscovery Biomarkers  
Organized by: GTCbio  
The Boston Park Plaza Hotel & Tower  
Boston, MA

**May 19-21, 2008**

GOT Summit  
Organized by: Cambridge Healthtech Institute  
Hilton Boston Logan Airport  
Boston, MA

**May 9-10, 2008**

Translating Proteomic Discoveries into Clinical Diagnostics  
Organized by: American Association for Clinical Chemistry  
Grand Hyatt Hotel  
Seattle, WA

**March 14-15, 2008**

Affinity Reagents for Human Proteins  
Elite Palace Hotel Stockholm  
Stockholm, Sweden

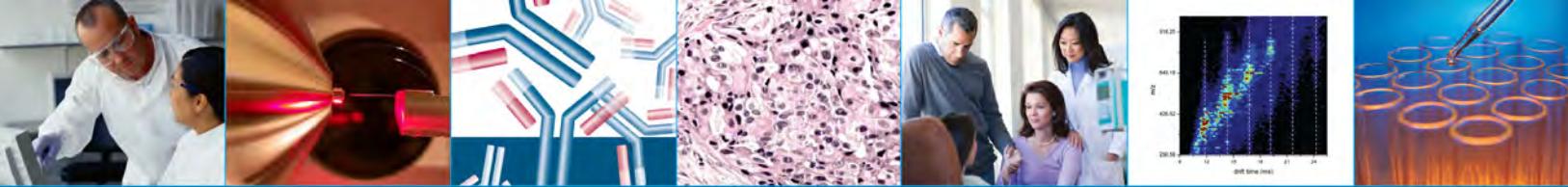
**January 14-15, 2008**

Oncology Biomarkers  
Organized by: GTCbio  
Hotel Whitcomb  
San Francisco, CA

**January 9-11, 2008**

Mitochondria Minisymposium 2008  
Natcher Conference Center, National Institutes of Health  
Bethesda, MD

2007



**November 15, 2007**

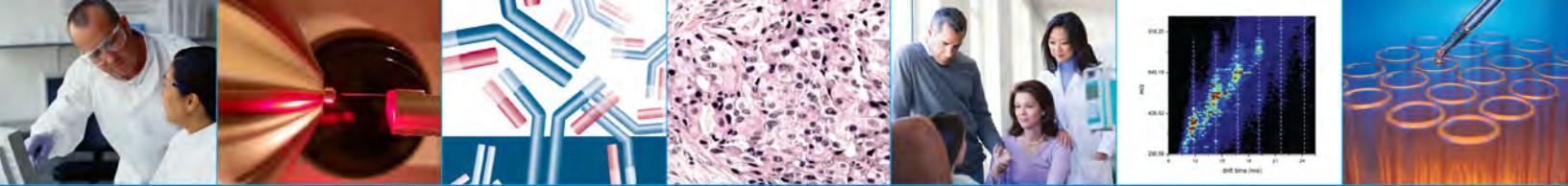
Strategies for Improving Reliability in Protein/Peptide Identification Workshop  
National Institute of Standards and Technology Campus  
Gaithersburg, MD

**October 24-25, 2007**

Clinical Proteomic Technologies Initiative for Cancer Annual Meeting  
Hilton Washington DC / Rockville  
Rockville, MD

**September 24-25, 2007**

Florida International University Presentation  
Organized by: Florida International University  
Miami, FL



## 4.16 CPTC Program Staff Biographies

**Anna D. Barker, Ph.D.**

**Deputy Director, Advanced Technologies and Strategic Partnerships  
National Cancer Institute**

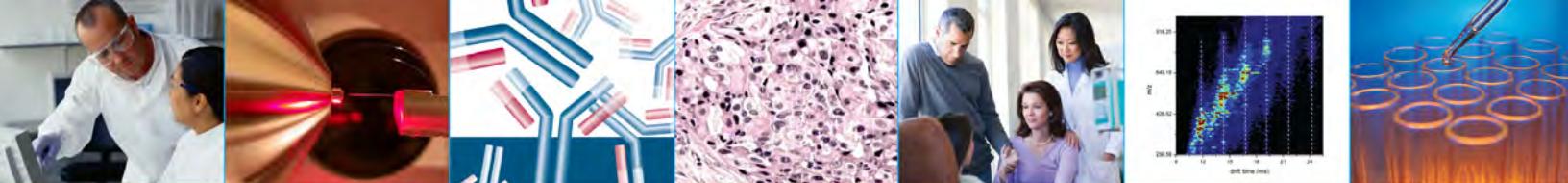
Dr. Barker serves as the Deputy Director for Advanced Technologies and Strategic Partnerships of the NCI. In this role she plans and coordinates the implementation of integrative, multi-disciplinary and multi-sector programs to accelerate the development and translation of new knowledge and advanced technologies into effective interventions to prevent, detect and treat cancer. Under her leadership, the NCI has recently launched programs in bioinformatics and nanotechnology and is currently developing initiatives in proteomics and biospecimens to enable cancer research. Dr. Barker completed her Ph.D. at the Ohio State University, where she trained in immunology and microbiology. Her research interests include experimental therapeutics, tumor immunology, and free-radical biochemistry in cancer etiology, prevention and treatment.

Dr. Barker has a long history in research and the leadership and management of research and development, technology transfer and product commercialization in the non-profit and private sectors. Prior to entering the biotechnology sector, she was a senior executive at Battelle Memorial Institute for 18 years where she developed and led a large group of scientists and technical staff working in areas such as drug discovery and development, pharmacology and biotechnology, including several NCI sponsored research programs. In the private sector she co-founded and served as the CEO of a public biotechnology company, focused in therapeutics discovery and development; and subsequently founded and served as the CEO of a private company dedicated to the transfer and deployment of technologies to prevent, diagnose and treat cancer.

She is a member of the Steering Committee of C-Change and Chairperson of the C-Change Cancer Research Team. She is a member of the DOD Breast Cancer Research Program Integration Panel, and a past chairperson of the BCRP Integration Panel. Dr. Barker has served in several capacities for the AACR, including the Board of Directors and chairperson of the Public Science Policy and Legislative Affairs Committee; and the NCI, including membership on the Board of Scientific Counselors for the Division of Cancer Etiology and chairperson of the Cancer Center Support Review Study Section. Dr. Barker has received a number of awards for her contributions to research, cancer patients, professional and advocacy organizations and the ongoing national effort to prevent and cure cancer, including a named fellowship from the AACR.

**Henry Rodriguez, Ph.D., M.B.A.**

**Director, Clinical Proteomic Technologies for Cancer  
Office of the Director  
National Cancer Institute**



Dr. Rodriguez serves as the Director of the Clinical Proteomic Technologies for Cancer programs within the NCI's Office of Technology and Industrial Relations, in the Office of the Director. In this role, he oversees and is responsible for the vision, direction, and implementation of highly complex scientific proteomic technology programs. These initiatives consist of the Clinical Proteomic Technologies for Cancer initiative (CPTC) and the Mouse Proteomic Technologies Initiative (MPTI).

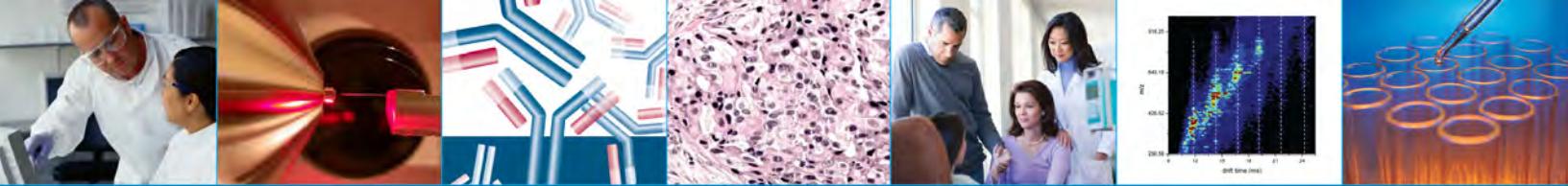
Dr. Rodriguez is an internationally recognized expert in advanced molecular-based cancer technologies, specifically proteomics, genomics, cellomics, and bioinformatics. Immediately before coming to the NCI in 2006, Dr. Rodriguez was at the National Institute of Standards and Technology (NIST), where he held several roles (1998-2006). At NIST, Dr. Rodriguez developed and was the Leader of the Cell & Tissue Measurements Group, where he successfully established four advanced technology research programs in the areas of quantitative cell biology/cellomics, proteomics, gene expression, and bioinformatics. Initiatives in proteomic metrology included protein capture chemistry, mass spectrometry, protein sequencing, peptide synthesis, and database development.

Dr. Rodriguez also developed a tissue engineering program, of which he was the program manager and chairman of the NIST Regenerative Medicine Strategic Working Panel. In this capacity, he established molecular-based analytical procedures to measure genetic damage in tissue-engineered medical products that may result from manufacturing, storage or shipping, and coordinated these efforts with the US Food and Drug Administration (FDA) and ASTM International.

Dr. Rodriguez helped pioneer new measurement methodologies that utilized liquid chromatography/mass spectrometry (LC/MS). He is internationally known for his outstanding efforts in developing, applying, and quantifying oxidatively-modified base detection methods for applications in health care and cancer diagnostics. Dr. Rodriguez has been awarded several Department of Commerce Advanced Technology Program (ATP) intramural awards and served as an ATP Proposal Scientific Expert.

Dr. Rodriguez served as a Program Analyst (Science/Strategic Policy Expert) in the Office of the Director at NIST, where he helped in the planning and evaluation of strategic scientific policies and partnerships in biotechnology, including interactions with the Department of Commerce and members of Congress. In 2005, Dr. Rodriguez developed the *Measurement Challenges in Proteomics* workshop, as part of NIST's Roadmap Initiative. He also participated on Department of Commerce efforts in education and/or training in genetics and genetic technologies of professionals for the Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS).

Dr. Rodriguez is an advocate that molecular medicine is transforming the processes of drug discovery, development, and delivery, suggesting that the war on cancer is embarking upon a new era. By providing a suite of rapid, sensitive, specific, high-throughput analytic tools and supporting standardized reagents and informatics



capabilities, the integrated application of advanced technologies offers the possibility of transforming this challenge into an opportunity.

Dr. Rodriguez began his career as a fellow at the Department of Immunology of The Scripps Research Institute, and then at the Department of Medical Oncology of the City of Hope National Medical Center. Dr. Rodriguez is the recipient of domestic and international awards, including the Sigma Xi Young Scientific Investigator award, the Alumni Honor Roll of the MARC/MBRS programs of the National Institute of General Medical Sciences at the National Institutes of Health, a Proclamation letter signed by the Governor of Hawaii, and the Science Spectrum Trailblazer, Top Minority in Research Science Award. In addition, he has been elected to the presidency of scientific organizations such as the Oxygen Club of Greater Washington, D.C. He has authored over 59 papers in peer-reviewed journals and books and co-edited a science book titled Oxidative Stress and Aging: Advances In Basic Science, Diagnostics and Intervention, that made the publisher's "Bestsellers List."

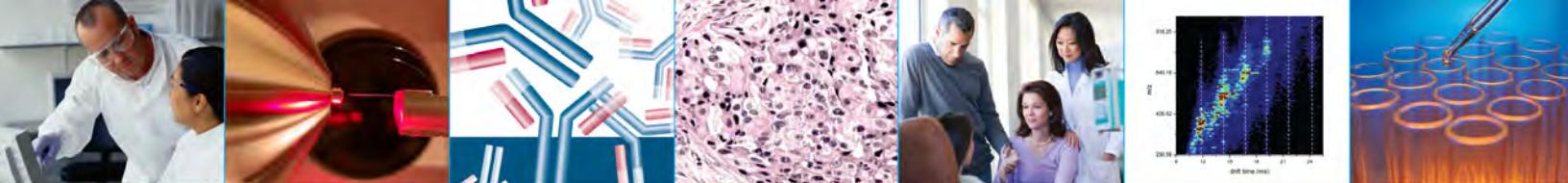
Dr. Rodriguez received a B.S. and M.S. in Biology from Florida International University, a Ph.D. in Cell and Molecular Biology from Boston University, and an M.B.A. from Johns Hopkins University School of Business Management.

**Tara Hiltke, Ph.D.**  
**Program Manager, Clinical Proteomic Technologies for Cancer**  
**Office of Technology and Industrial Relations**  
**National Cancer Institute**

Dr. Tara Hiltke serves as Program Manager for the Clinical Proteomics Technologies for Cancer initiative, a program in the NCI Office of Technology and Industrial Relations. In this role she serves as a technical resource for the integration of proteomic reagents and resources into targeted diagnostics and therapeutics.

Dr. Hiltke has nearly ten years of post graduate experience in immunology, biochemistry and molecular biology. Prior to NCI, she served as a senior scientist/project manager in assay development at both Wellstat Diagnostics and BioVeris Corporation. She developed clinical assays for diagnostic markers through the use of an electrochemiluminescence platform and magnetic beads. Prior to that, she led a number of projects at GenCyte LLC which focused on a mammary associated chemokine. Specifically she investigated the diagnostic and therapeutic potential of this marker through the use of ELISA and animal models respectively.

Dr. Hiltke received a B.S. from Nazareth College, where she was in the college honors program and majored in Biology/minored in Chemistry. She obtained her Ph.D. from the Center of Advanced Molecular Biology and Immunology (CAMBI) at the University of Buffalo in 1999.



**Mehdi Mesri, Ph.D.**

**Program Manager, Clinical Proteomic Technologies for Cancer**

**Office of the Director**

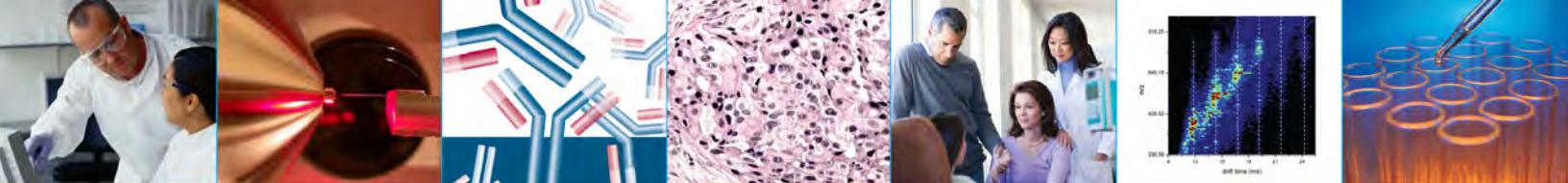
**National Cancer Institute**

Dr. Mesri serves as program manager of the Clinical Proteomic Technologies for Cancer initiative, a program in the NCI's Office of Technology and Industrial Relations. In this role, he coordinates novel proteomics and genomics technology development, assessment, and metrology development projects for the program. In addition, Dr. Mesri serves as a technical resource on integrating emerging technologies in proteomics, genomics, and pharmacogenomics into targeted diagnostics and therapeutics. Dr. Mesri performs scientific analysis and coordinates the implementation of trans-divisional cancer proteomic technology research initiatives.

Dr. Mesri has fourteen years of post-PhD experience including seven years of biotechnology/biopharmaceutical industry experience in drug discovery/validation in cancer, vascular biology, and immunology. Prior to the NCI, Dr. Mesri served as a principal scientist/projects manager in the Department of Protein Therapeutics at Celera. There, he used mass spectrometry technologies to discover and validate biologic antibody targets in oncology, including prostate cancer and angiogenesis. Using in vitro and in vivo models, he has contributed to the validation of biologic and small molecule targets using validation platforms including proteomics, antibody and/or toxin-conjugated antibody-mediated cell death, RNAi, IHC, FACS, and apoptosis. Dr. Mesri was also instrumental in a lung cancer biomarker/diagnostic discovery and validation program at Celera, which was aimed at generation of a constellation of biomarkers for early detection and management of lung cancer. Prior to that Dr. Mesri was a project leader at CuraGen Corporation, where he championed genomically-derived high value protein drug and fully human monoclonal antibody targets for an immunology/inflammation program, including a new protein with a novel immunosuppressive function which he advanced to the pre-clinical stage. There he also established a functional screening assay panel, developed bio-analytical ELISA methods for CuraGen clinical candidates, presented at the IND team meetings, and established/managed external collaborations on target candidates, among other responsibilities.

During his five-year postdoctoral training, and later as a junior faculty at Yale University, Dr. Mesri led a number of projects. He investigated cancer gene therapy by developing a survivin (member of inhibitors of apoptosis, IAP) mutant adenovirus in xenograft breast cancer models. He was interested in trying to bridge the survivin story with some aspects of vascular biology and he pursued some initial observations and demonstrated that survivin was actually a key cell viability gene downstream of VEGF stimulation and supported the data using animal studies.

Dr. Mesri is the recipient of domestic and international awards including the CuraGen Corporation Special Achievement Award, a Reciprocal Exchange Fellowship by the British Society for Immunology, a federally funded National Research Service Award, and a Scottish Medical Research Studentship.



Dr. Mesri received a B.Sc. in Biomedical Sciences from the University of Bradford, (UK), a M.Med.Sci. in Clinical Pathology from the University of Sheffield (UK), and a Ph.D. in Immunology from the University of Aberdeen (UK).

**Christopher Kinsinger, Ph.D.**

**Program Specialist, Clinical Proteomic Technologies for Cancer**

**Office of the Director**

**National Cancer Institute**

Dr. Kinsinger serves as program specialist of the Clinical Proteomic Technologies for Cancer initiative, a program in the NCI's Office of Technology and Industrial Relations. In this role, he oversees the development and assessment of technology platforms for cancer research. Dr. Kinsinger focuses on programmatic goals involving mass spectrometry, informatics, and biospecimens. In this role he works with NCI staff and investigators to optimize proteomics technology, establish policies for sharing data and biospecimens, and generally improve the quality and reliability of proteomic data and samples. Dr. Kinsinger completed his B.S. degree in 1999 from Wheaton College and Ph.D. degree in chemistry in 2004 from the University of Minnesota. Prior to joining the NCI's Clinical Proteomics Management team, Dr. Kinsinger completed postdoctoral training at the National Institute of Standards and Technology, where he researched fragmentation pathways of peptide ions in mass spectrometry.