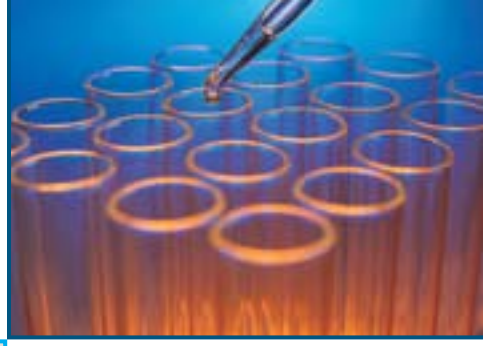
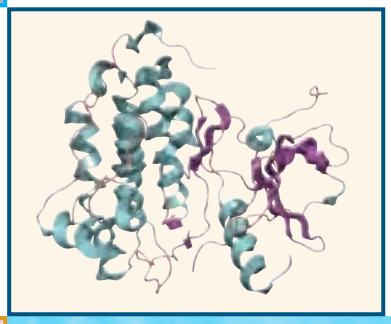
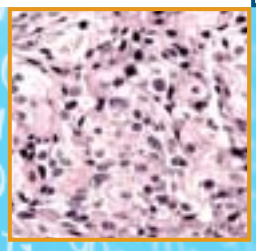
National Cancer Institute

# CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER







2008-09 | ANNUAL REPORT

# *Advancing Protein Science for Personalized Medicine*

## Building a Reliable and Efficient Protein Biomarker Pipeline

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

# TABLE OF CONTENTS

# EXECUTIVE SUMMARY

The addition of protein biomarker panels to the cancer diagnostic toolkit is an area of considerable interest in medicine. The discovery that proteins and peptides are "leaked" by tumors into clinically accessible body 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.

> "Proteomics is going to be an enterprise that will go on for decades. In many ways executing the Human Genome Project was very simple by comparison."
>
> Francis Collins, M.D., Ph.D.
> Director, National Institutes of Health

Prostate-specific antigen (PSA) and cancer antigen 125 (CA-125) are examples of blood-borne cancer protein biomarkers that are currently 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 proteins are under intense investigation since biomarker panels will potentially bring greater sensitivity and specificity to cancer screening than any one biomarker 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. There are well over 1,000 candidates 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. This discrepancy indicates that the issue lies within the candidate biomarker pipeline.

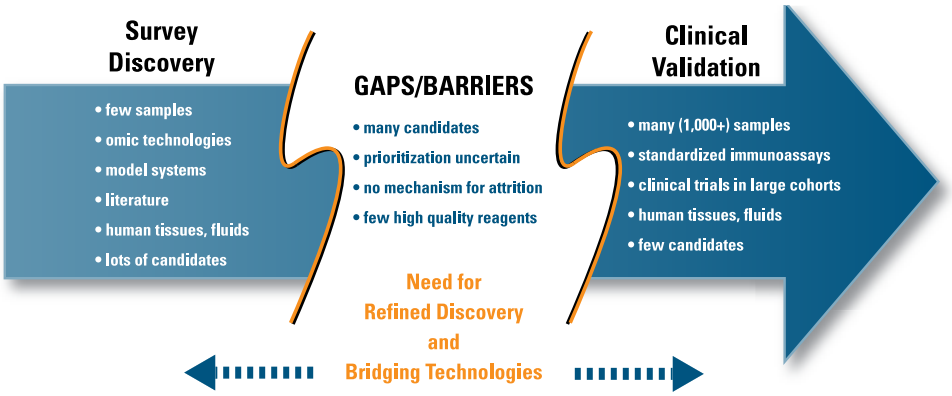
Proteomic technologies hold great promise for the discovery of novel cancer biomarkers; however, studies that have applied proteomic technologies to clinical applications have met with some disappointment. There are two major issues at hand:

• *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.

• *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, each ELISA may take up to one year and millions of dollars to develop. A more efficient biomarker development pipeline will require technology that bridges discovery and clinical validation in order 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 should rapidly triage a lengthy list of candidates prior to investing very large sums of time and money on the development of antibodies suitable for use in an ELISA (Figure 1).

Traditional approaches have contributed all they can; it is time for new technologies and new 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.

**FIGURE 1. A Better Bridge Is Needed Between Biomarker Discovery and Clinical Validation**




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
- 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.)
- Making high-quality reagents available and accessible
- Developing technologies that can quantify proteins across a large dynamic range
- Developing bioinformatics resources with shared algorithms and standards for processing, analyzing, and storing proteomic data
- Implementing a verification step in the protein biomarker pipeline
- Adopting an interdisciplinary team approach to science

In many ways, the challenges facing the clinical proteomic community are comparable to those that once faced 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 might never have been able to live up to its promise, and the opportunities afforded by the HGP might never have been realized.

> "CPTC has made great strides in bridging the gap between laboratory advances and clinical utility of proteins by developing a pipeline with greater accuracy through the process of pre-clinical verification."
>
> John E. Niederhuber, M.D.
> Director, National Cancer Institute

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 of 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 direct impact is the ultimate reason we are all tirelessly working to ensure the success of this program.

CPTC Program Coordinating Committee

## CPTC by the Numbers

**4** Partnerships with federal agencies and professional organizations

**6** Datasets publically available

**7** Patents

**11** Partnerships with biotechnology companies

**12** Leveraged funding activities

**19** Partnerships with academic institutions

**26** Software tools

**27** Standard Operating Procedures

**>60** Number of organizations that make up the CPTC Community

**84** Monoclonal antibodies

**171** Publications*

*CPTC was launched in 10/2006 and awarded 20 grants (5 U24, 4 R21, 3 R21/R33, and 8 R01). RFAs were non-reissue. As of 10/2009, 15 grants encompass the CPTC portfolio (5 U24, 1 R21, 2 R33, and 7 R01).

# TEAM SCIENCE: THE CPTC COMMUNITY

Solving big problems requires big ideas and strong teams for execution. Removing the barriers to the application of proteomic technologies to clinical cancer research is a big problem that must be overcome with big ideas and strong teams if we hope to detect cancer at its earliest stages, allowing us to treat the disease before it can exert its devastating personal and societal effects.

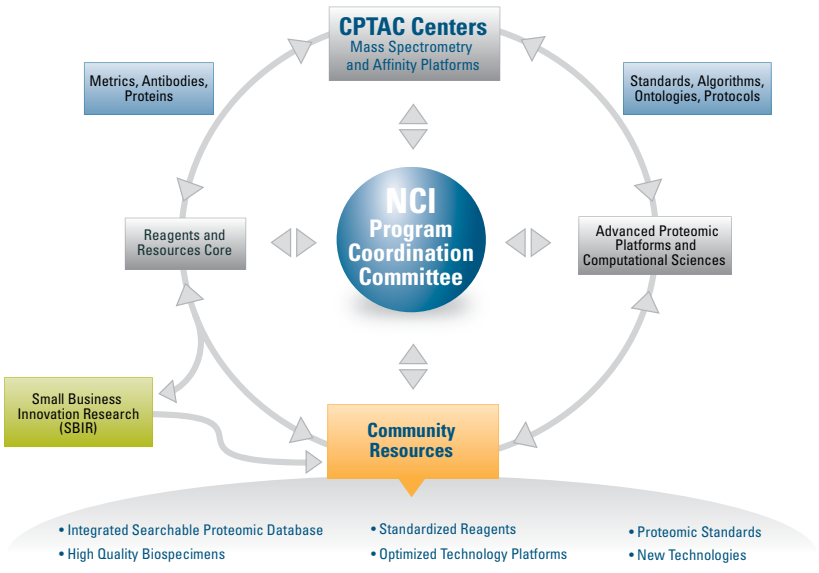
The NCI Clinical Proteomic Technologies for Cancer (CPTC) is a five-year initiative composed of three integrated components designed to remove the barriers to the clinical application of proteomic technologies (Figure 2). No one laboratory working on its own could possibly achieve all of the goals necessary for the success of such a lofty goal. It takes an interdisciplinary team science approach (Figure 3).

The CPTC community consists of scientists from nearly 50 federal, academic, and private sector organizations, which include seasoned and senior investigators who are leading large centers and multi-project efforts as well as junior faculty involved in individual projects.

**FIGURE 2. CPTC Program Components**

| CPTAC (Clinical Proteomic Technology Assessment for Cancer) | Advanced Proteomic Platforms and Computational Sciences | Proteomic Reagents and Resources Core |
|---|---|---|
| Optimizes existing proteomic platforms to reliably identify, quantify, and compare peptides/proteins in complex biological mixtures through a multidisciplinary network of centers | Supports development of highly innovative research in the "quantitative" analysis of peptides/proteins of interest in clinical cancer studies | Provides central source of proteins, peptides, antibodies, proficiency testing materials, and informatics tools |

**FIGURE 3. Team Science**



# Clinical Proteomic Technology Assessment for Cancer (CPTAC) Network

> 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 and quantitatively tested. The advances made by CPTAC investigators in the last two years are very encouraging."
>
> Leland Hartwell, Ph.D.
> President and Director
> Fred Hutchinson Cancer Research Center
> Recipient of the 2001 Nobel Prize in Physiology or Medicine

Clinical proteomics research is hampered by a lack of standardized technologies and methodologies, which are critically needed in order to discover and validate proteins and peptides relevant to cancer, also known as biomarkers, more effectively. To address this critical need, the NCI established a collaborative network of five Clinical Proteomic Technology Assessment for Cancer (CPTAC) teams.

The CPTAC network extends well beyond these five centers, however, bringing in expertise from both the public and private sectors to create one of the most in-depth, multidisciplinary networks focused on realizing the potential of clinical proteomics (Figure 4). The network's ultimate goal is to enable all researchers conducting cancer protein biomarker research at different laboratories to use proteomic technologies and methodologies to directly compare and analyze their work. This network should lead, in turn, to improved diagnostics, therapies, and even prevention of cancer.

**FIGURE 4. The CPTAC Network**

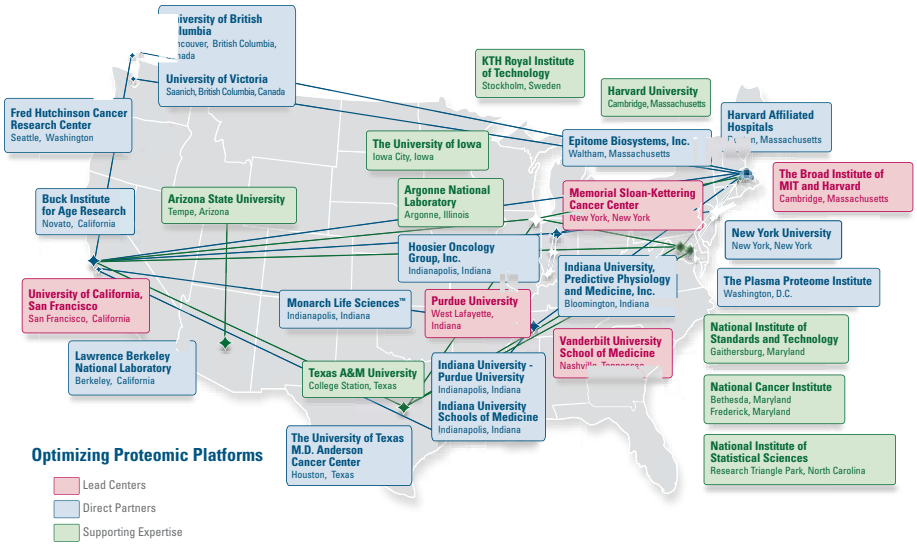


**FIGURE 5. New Protein Biomarker Development Pipeline**

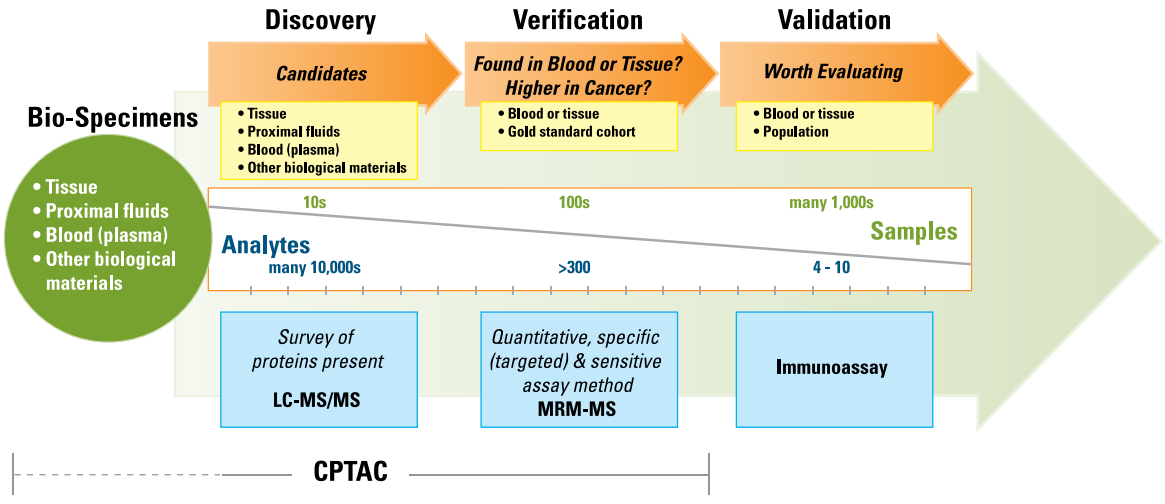

CPTAC investigators are in the process of developing an integrated biomarker development pipeline, a sort of "highway" from discovery to clinical application (Figure 5). In support of this endeavor, CPTAC has conducted a series of "round robin" studies designed to accomplish two over-arching goals (Figure 6):

- Refine protein biomarker discovery through metrics
- Introduce a pre-validation step, verification, into the biomarker development pipeline using quantitative multiplex "bridge" technologies

Through both intra- and inter-laboratory studies, CPTAC has made substantial advances in improving the process of biomarker discovery and understanding the appropriate structure for a real biomarker pipeline.

> "There are a multitude of variables that pose significant challenges to the translation of proteomic discoveries into clinical applications. The best way forward is through highly collaborative and integrated consortia of expert proteomic groups such as those formed through the NCI's CPTAC program."
>
> 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
> The University of Texas M.D. Anderson Cancer Center

**FIGURE 6. CPTAC is refining the discovery stage and introducing a pre clinical verification stage in the biomarker development pipeline for greater accuracy and efficiency**




*Adapted from Rifai, Gillette and Carr, Nat. Biotech, 2006*

## *Inter-Laboratory "Round Robin" Studies*

### Multiple Reaction Monitoring Performance Assessment

Candidate-based biomarker verification is increasingly being viewed as a critical step in the biomarker discovery pipeline, requiring quantitative, rapid, and sensitive assays to bridge the gap between unbiased discovery and pre-clinical validation. In the past, verification of novel biomarkers has relied on the use of high-throughput immunoassays. The development of these assays relies on the availability of suitable, well-characterized antibodies; however, such reagents for novel biomarkers do not exist currently, and the time, expense, and technical limitations required to generate them provide a strong incentive to develop alternative approaches.

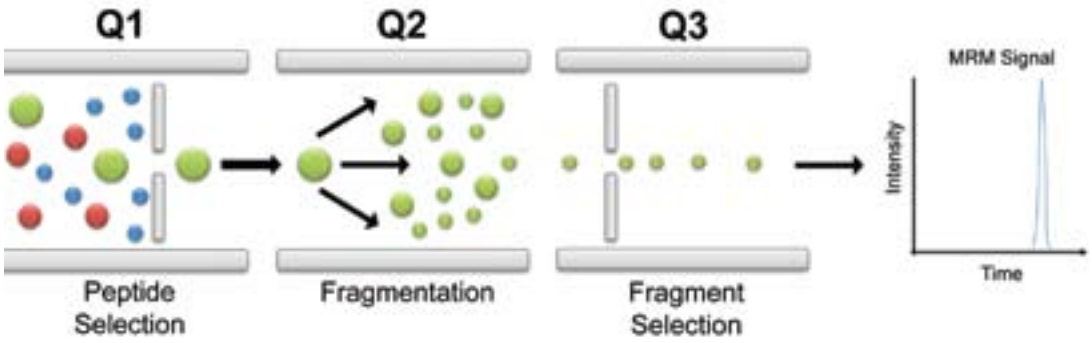
Unlike traditional mass spectrometry, which attempts to detect all proteins in a biological sample in a scattershot fashion, Multiple Reaction Monitoring (MRM) is highly selective (targeted), allowing researchers to fine tune an instrument to specifically look for peptides, or protein fragments, of interest. This approach allows for greater specificity, sensitivity, speed and quantitation of an analyte of interest (biomarker candidate). As a result, MRM has application for the verification of candidate biomarkers from discovery experiments, ensuring that highly qualified candidates move into clinical validation. CPTAC has been pioneering the use of this targeted mass spectrometry method in its biomarker development pipeline.

Recently, several individual laboratories have successfully used targeted multiple reaction monitoring (MRM) assays coupled with stable isotope dilution mass spectrometry (SID-MS) to quantify proteins in human plasma directly (Figure 7). Using this method, it becomes feasible to rapidly screen large biospecimen banks in pre-clinical studies to obtain statistically valid data for candidate biomarker 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 the intra- and inter-laboratory analytical performance of SID-MRM-MS assays. The results of these studies include the following highlights:

- An MRM 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.
- The reproducibility of MRM protein quantitation across multiple sites with differing liquid chromatography–mass spectrometry (LC-MS/MS) platforms has been demonstrated.

**FIGURE 7. Scheme of a Multiple Reaction Monitoring (MRM) experiment. The first mass analyzer (Q1) is set to only transmit the parent weight of a protein, the collision energy is optimized to produce a diagnostic charged fragment of a protein fragment (peptide) in the second mass analyzer (Q2), and the third mass analyzer (Q3) is set to transmit this diagnostic peptide fragment only. Therefore, only this exact peptide transition is detected.**



- The greatest source of MRM variability lies with sample processing.

Through these experiments, CPTAC has demonstrated that SID-MRM-MS technology can be used successfully to bridge the gap between unbiased discovery and pre-clinical validation in a reproducible and repeatable way.

PUBLICATION:
Addona, TA, *et al.* **A Multi-Site Assessment of Precision and Reproducibility of Multiple Reaction Monitoring-Based Measurements: Toward Quantitative Protein Biomarker Verification in Human Plasma**. *Nature Biotechnology* 2009; 27(7):633-641.

■ ■ ■

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

The lack of a widely available reference standard as a quality control sample for benchmarking the performance of liquid chromatography-tandem mass spectrometry (LC-MS/MS) platforms has hindered the proteomics community. As a result, CPTAC has developed a well-characterized yeast *Saccharomyces cerevisiae* reference proteome as a resource for the community. This reference standard will make it easier to standardize both current and emerging proteomic technologies and help ensure that only the highest quality data are being generated.

A CPTAC network study has described a large-scale production of this yeast reference proteome, which the National Institute of Standards and Technology (NIST, www.nist.gov) now offers. Furthermore, a historic reference dataset has been created that characterizes this reference proteome and can be used to define the performance of individual ion trap–based LC-MS/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/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.

PUBLICATION:
Paulovich AG, *et al.* **A CPTAC Inter-Laboratory Study Characterizing a Yeast Performance Standard for Benchmarking LC-MS Platform Performance.** *(2009) Molecular & Cellular Proteomics.* In Press.

■ ■ ■

## 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. This method is used increasingly 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 reflect the underlying biology of disease phenotypes accurately. 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 made during an experiment. While these numbers can indicate variability, they cannot indicate whether system performance is optimal or show which components require optimization. A CPTAC consortium-wide study has described 44 metrics for evaluating the performance of LC-MS/MS system components, and the study implements 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 for those across laboratories.

PUBLICATION:
Rudnick, PA, *et al.* **Performance Metrics for Liquid Chromatography-Tandem Mass Spectrometry Systems in Proteomic Analyses and Evaluation by the CPTAC Network.** *(2009) Molecular and Cellular Proteomics.* In press.

■ ■ ■

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

Repeatability and reproducibility in LC-MS/MS both give rise to the variability observed in analytical proteomics, which is widely perceived as problematic but has never been evaluated systematically. Thus, the field remains unaware of their limits and the degree to which standardization can be achieved.

To evaluate measures of variability, CPTAC network institutions contributed data from 12 instruments for three different samples, each of which differed in protein number and dynamic range. Implementing a Standard Operating Procedure (SOP) resulted in a reduced number of peptide identifications. This reduction suggests that diverse protocols for LC-MS/MS systems enable researchers to identify a

## CPTAC, BIOSPECIMENS, AND NCI BEST PRACTICES

OBBR Office of Biorepositories and Biospecimen Research

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

At the launch of the CPTAC Network, the centers quickly realized that each site 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 (Massachusetts General Hospital/Harvard Medical School) 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 Standard Operating Procedure 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 multicenter, multidisciplinary projects.

In addition, CPTC's program staff serve as proteomics scientific technical experts to OBBR's Biospecimen Research Network.

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 rigorous approach is needed. Using SOP-driven methods, LC-MS/MS proteomics can distinguish proteomic differences amounting to less than 10 percent of the detectable proteome. Therefore, the study found that carefully adhering to SOP-driven methods could effectively standardize LC-MS/MS proteomics across laboratories, optimizing biomarker discovery.

PUBLICATION:

Tabb, DL, *et al.* **Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography - Tandem Mass Spectrometry.** *(2009) Journal of Proteome Research*. In Press.

## *Intra-Laboratory Studies*

### The Broad Institute of MIT and Harvard

PRINCIPAL INVESTIGATORS:

**Steven A Carr**, Ph.D., The Broad Institute of MIT and Harvard

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

**Leigh Anderson**, Ph.D., Plasma Proteome Institute

OTHER PARTICIPATING INSTITUTIONS:

- Dana-Farber Cancer Research Institute
- Epitome Biosystems. Inc.
- Epitomics, Inc.
- Massachusetts General Hospital
- University of Victoria
- University of Victoria Genome BC Proteomics Center
- University of Washington

RESEARCH INTERESTS:

- Multiple reaction monitoring (MRM) assays for quantification
- SISCAPA (sable isotope standards and capture by anti-peptide antibodies)

The absence of methods capable of assessing large numbers of protein biomarker "candidates" emerging from discovery "omics" experiments is the primary bottleneck impeding the development of improved cancer diagnostics. The overarching goal of the Broad Institute CPTAC effort is to eliminate this bottleneck by developing a sensitive, specific, and quantitative technology based on multiple reaction monitoring mass spectrometry (MRM-MS) that is capable of measuring hundreds of candidate cancer biomarker proteins in large sets of clinical plasma samples.

The team is developing various methods that enable protein measurement at concentrations in plasma 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 enrichment. The researchers are developing SISCAPA-MRM assays for more than 100 candidate biomarker proteins and are using them to measure candidate protein levels in plasma from breast cancer patients.

The team has increased the sensitivity of MRM multiplexed assays for proteins in plasma by more than 500-fold and has demonstrated, for the first time, that multiplexed MRM-MS-based assays can be constructed reliably and applied robustly to quantify biomarker candidates that are present in plasma in tiny quantities (low ng/mL range). More than 20 SISCAPA-MRM assays have been constructed, including the first 9- and 10-plex assays, demonstrating the real power of SISCAPA for clinically relevant proteins and its ability to multiplex.

The team has established an optimized antibody reagent production and quality control pipeline to support the creation of SISCAPA assays for candidate cancer biomarkers. So far, this pipeline has been used to generate reagents for more than 100 CPTAC candidate proteins, as well as 16 other proteins targeted in four separately funded biomarker

## EMPOWERING THE SCIENTIFIC COMMUNITY WITH QUANTITATIVE TOOLS TO MEASURE ALL HUMAN PROTEINS *FEASIBILITY STUDY LAUNCHED*

The next frontier in personalized medicine, analyzing the human proteome, might be getting a bit closer, thanks to federal stimulus funding from the National Cancer Institute. The grant was awarded to two members of the CPTAC Network: Amanda Paulovich, M.D., Ph.D., Fred Hutchinson Cancer Research Center, and Steven Carr, Ph.D., The Broad Institute of MIT and Harvard. Together, they will co-lead a pilot study that will assess the feasibility and scalability of a method for quantitatively measuring all of the proteins in the human body. The long-term output of the project will hopefully be the human Proteome Detection and Quantitation (hPDQ) project.

"Currently the biomedical research enterprise is severely hindered by its inability to measure the vast majority of human proteins – and you can't study what you can't measure," says Paulovich.

This study is designed to change that. Paulovich and colleagues will use a highly sensitive emerging technology based on multiple reaction monitoring mass spectrometry (MRM-MS) to develop assays that will measure the levels of 200 proteins found in breast cancer cells. The strength of this technology is that it will enable researchers to develop multiplexed assays, which can measure large numbers of proteins in complex biological specimens simultaneously.

"This pilot has the potential of developing the first step toward making the entire human proteome clinically accessible," said Henry Rodriguez, Ph.D., M.B.A. director, CPTC. If successful, this pilot could have two powerful outcomes: stimulating a larger international endeavor to assess the utility of all human proteins as potential biomarkers in hundreds of diseases in a very efficient manner, and secondly, empowering the global research community with necessary tools and reagents to better study biological processes.

"If we can create ways to measure a large fraction of human proteins, particularly those in very low abundance, this will facilitate the development of new drugs and personalized medicine," Paulovich said.

projects. Peptide capture, wash, and elution on antibody-coated magnetic beads for SISCAPA have also been optimized and automated.

The team has also developed several computational tools, including an algorithm to prioritize lists of cancer-associated proteins for assay development, an algorithm for selecting the best peptides from candidate proteins to use for MRM and SISCAPA-MRM assays, and an Accurate Inclusion Mass Screening (AIMS) tool for qualifying hundreds of protein candidates per week in plasma prior to committing to the resource intensive steps of establishing a quantitative assay.

■ ■ ■

## Memorial Sloan-Kettering Cancer Center (MSKCC)

PRINCIPAL INVESTIGATOR:

**Paul Tempst**, Ph.D.

OTHER PARTICIPATING INSTITUTIONS:

- New York University Langone Medical Center

RESEARCH INTERESTS:

- Sample fractionation using magnetic beads for capture of peptides
- Automated sample processing technology (robotics)
- Standardized approaches to proteolytic activities

Since 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 CPTAC team is evaluating and documenting whether serum peptide patterns, and/or the protease activities producing them, can be measured reproducibly and whether they have diagnostic value for cancer detection, mark a given clinical outcome, or can distinguish between clinically significant and insignificant cancers.

The MSKCC team developed an exopeptidase (enzymes that remove the first or last amino acid from a polypeptide chain) activity-based test that is capable of measuring global exopeptidase activities within individual proteomes of two or more groups of biological fluids. This test uses semi-automated Matrix Assisted Laser Desorption/Ionization-Time of Flight (MALDI-TOF) mass spectrometry, labeled peptide substrates, and non-degradable reference peptide internal standards. Synthetic substrates and non-degradable reference peptides have been produced and tested in serum and plasma exopeptidase assays and have been made available to the scientific community. Advantages of this approach over standard peptidome measurements are robustness, reproducibility, and quantitation.

The MSKCC team has also made significant improvements on the reproducibility and repeatability of MALDI-TOF mass spectrometry. Initial comparisons of MALDI-TOF reproducibility between the MSKCC prototype setup and the New York University mirror site for serum (endogenous) 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 toward similar outputs. The team has since made major improvements on aligning peaks along the x-axis (m/z) and normalizing ion intensities (y-axis); however, these observations put practical limitations on the selection of diagnostic peak patterns. Similar comparative assessment of the assay output patterns is underway.

■ ■ ■

## Purdue University

PRINCIPAL INVESTIGATOR:

**Fred E. Regnier**, Ph.D.

OTHER PARTICIPATING INSTITUTIONS:

- Hoosier Oncology Group
- Indiana University Purdue-University at Indianapolis
- Indiana University School of Medicine
- Indiana University
- Predictive Physiology and Medicine, Inc.
- Monarch LifeSciences

RESEARCH INTERESTS:

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

MS-based proteomic 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 because of the complexity of human biofluids and the protocols employed in these approaches.

To help overcome this limitation, the Purdue CPTAC team has developed robust protocols and standards for both electrospray ionization (ESI) and MALDI MS platforms, which include new or improved separation and/or enrichment systems.

The Purdue team designed and implemented a clinical proteomics data model for managing metadata of MS experiments. The data management system can be used to capture and manage experimental designs, experimental protocols, data storage parameters, search software, and annotation of result descriptions and has been implemented as a customized application of the metadata framework in the Computational Proteomics Analysis System (CPAS).

One of the team efforts toward standardizing proteomic data analysis among multiple laboratories has been to create a peptide annotation database for the proteomics community where biomedical researchers can search for plasma proteins collected from different mass spectrometry platforms together with experimental protocols and protein identification software for healthy individuals. In an effort to achieve this level of standardization, the Purdue team designed the Healthy Human Individual's Integrated Plasma Proteome (HIP2) Database. The primary goal of the HIP2 database is to support future clinical proteomics research, especially the discovery of biomarkers through plasma proteomics profiling.

A number of computational tools have been developed, including the development of a hierarchical statistical model to assess the confidence of peptide and protein identifications made by tandem mass spectrometry and the development of a Bayesian model for the protein inference problem in shotgun proteomics. The team also developed a Web-based Proteome Discovery Pipeline (PDP) for sharing proteomic data analysis tools among multiple laboratories. This Web-based analysis platform provides complete proteomics data analysis without requiring specialized hardware or input from bioinformatics specialists.

The Purdue team applied a label-free LC-MS/MS-based protein quantification technology to analyze plasma samples from 40 healthy women and 40 breast cancer (stage I & II) patients. The team identified 254 statistically significant differentially expressed proteins, of 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.

■ ■ ■

### University of California, San Francisco

PRINCIPAL INVESTIGATOR:

Susan Fisher, Ph.D.

OTHER PARTICIPATING INSTITUTIONS:

- Buck Institute for Age Research
- California Pacific Medical Center
- Lawrence Berkeley National Laboratory
- The University of Texas M.D. Anderson Cancer Center
- University of British Columbia

RESEARCH INTERESTS:

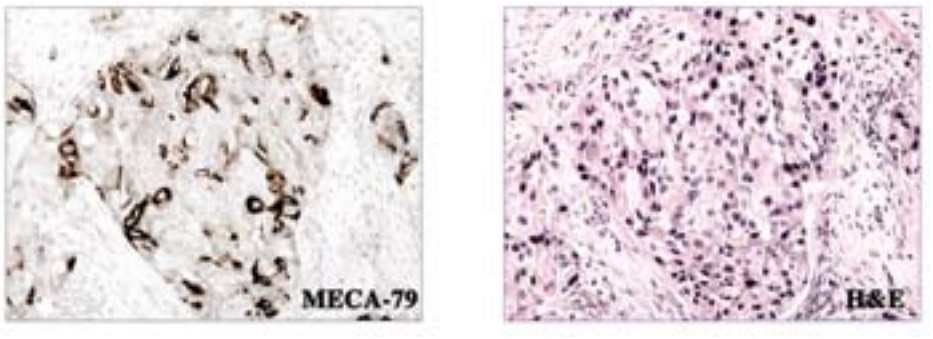
- Exploiting post-translational modifications (PTMs) for affinity capture of putative protein cancer biomarkers
- Identifying alternatively spliced breast cancer transcripts that can be used for early detection

The UCSF CPTAC team is evaluating proteomic technologies that will enable the early detection of several tumor types through the application of blood-based tests. The team is examining both global strategies and targeted MS-based approaches to develop optimal workflows for identifying protein signatures of human breast cancer cells. They are also establishing a systematic way to standardize proteomic protocols and data analysis.

Considerable progress has been made in developing and implementing affinity-based workflows for detecting proteins that carry various cancer-related PTMs, including glycoproteins, phosphoproteins, and modifications that are indicative of oxidative damage. To date, workflows for glycoproteins and phosphoproteins have been the major focus, and the team is close to launching a large inter-laboratory "round robin" CPTAC experiment targeting glycopeptides.

The group is also focusing on protein biomarkers that are specific for metastasis-prone (basal) subtypes, which are likely to have the largest impact on breast cancer survivorship. Approximately 1,700 genes have been identified that show strong evidence of alternative splicing across several breast cancer cell lines. In addition, UCSF investigators have shown that basal subtype cell lines and breast cancer tissues express a glycosylated protein variant that may be involved in the metastatic process (Figure 8). Identifying such candidate breast cancer biomarkers can potentially be used for early detection.

**FIGURE 8. Staining of tissues shows that a carbohydrate ligand with unusual functions in the immune system is expressed by human invasive (basal subtype) breast cancer cells (left panel). The UCSF team believes that this glycan may play a role in metastasis. H&E, hematoxylin and eosin stain, routinely used for tissue examination (right panel)**




## Vanderbilt University School of Medicine

PRINCIPAL INVESTIGATOR:
Daniel C. Liebler, Ph.D.

OTHER PARTICIPATING INSTITUTIONS:

- The University of Texas M.D. Anderson Cancer Center

RESEARCH INTERESTS:

- Extensive comparison of shotgun-based techniques
- Targeted measurements of biomarker proteins in tissues
- Reverse phase protein arrays

Shotgun proteomics is the most powerful platform that researchers use for analyzing complex proteomes. However, this approach has not been widely used to systematically detect proteomic differences between disease-specific phenotypes.

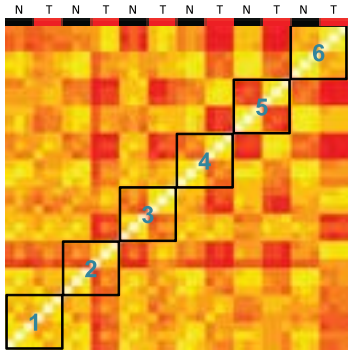
The Vanderbilt CPTAC team is trying to compare proteomes between cancer-relevant tissues to identify biomarkers. Although powerful, variable performance of the shotgun proteomics approach can prevent detection of very subtle differences between proteomes. To overcome this barrier, the Vanderbilt team has developed and implemented a new analytical platform for shotgun proteomics that is currently being applied to colon cancer samples.

Following successful proof of concept studies in colon cancer cell lines, the team is now using this new platform to analyze tissues retrospectively from patients with stage 2 colon cancers, with the goal of characterizing distinct protein biomarkers for colon cancer recurrence. If successful, their work could help pave the way for the next generation of cancer diagnostic testing.

The Vanderbilt team has developed a statistical model for distinguishing disease-associated phenotype differences from shotgun proteomics datasets. Application of this statistical model to proteomic datasets from pairs of colon tumor and normal tissue samples reveals global differences in protein expression that distinguish the tumors (Figure 9).

Other accomplishments of this team involve successfully evaluating shotgun proteomic analysis of formalin-fixed, paraffin-embedded (FFPE) tissues, developing and implementing an integrated, open-source data analysis pipeline for shotgun proteomics that exceeds the performance of other available tools, and developing a cost-effective method for targeted quantitation of hundreds of proteins, thereby providing an effective means to configure specific, targeted assays for biomarker candidates rapidly.

**FIGURE 9. Cluster map of shotgun proteome inventory comparison between colon adenocarcinomas and adjacent normal colon tissues**



# Advanced Proteomic Platforms and Computational Sciences

The Advanced Proteomic Platforms and Computational Sciences program is a comprehensive program focused on the development of innovative new tools, reagents, and the enabling of technologies for protein/peptide measurement. This program supports two 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 datasets

Investigators have made substantial achievements that are advancing the technological and analytical capabilities in proteomic research. These advances will allow the research community to better characterize and understand the differences between the normal and diseased human proteomes and to develop diagnostic and treatment procedures based on these distinctions.

■ ■ ■

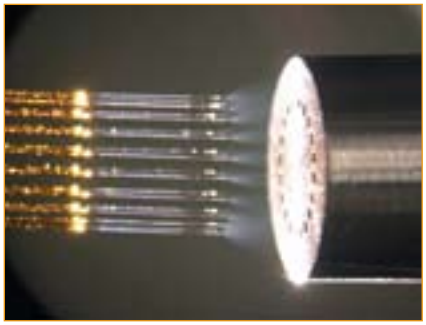
### A Proteomics Platform for Quantitative, Ultra-High-Throughput, and Ultra-sensitive Biomarker Discovery

INVESTIGATOR:

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

The Smith team is developing a next generation cancer biomarker discovery and validation platform to analyze clinically relevant samples that will provide measurements that are more robust, more sensitive, have higher throughput, and have improved quantitative utility in comparison to current platforms. This new platform combines fast liquid chromatography (LC) separations, ion mobility spectrometry (IMS), and time of flight (TOF) mass spectrometry (MS).

**FIGURE 10. Multi emitter ESI source provides 10-fold increase in ion production**



To enable more sensitive measurements using the LC-IMS-TOF MS systems, the team developed an advanced multi-emitter ESI source that provides more than 10-fold increase in ion production rates relative to the conventional single ESI emitter (Figure 10). In an evaluation of the platform, a complex tryptic digest of mouse plasma spiked with 20 reference peptides at varying concentrations 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 ≥ 100 ng/ml could be detected. In contrast, the LC-IMS-TOF MS platform was able to detect 19 of 20 spiked peptides at concentration levels down to 1 ng/ml.

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. In addition to providing a wider dynamic range for detection, the LC-IMS-MS system significantly reduced such interferences, allowing the team to detect many features that were undetectable using LC-MS. The LC-IMS-TOF MS platform achieves both sensitivity and throughput levels that are approximately one order of magnitude greater than those previously achieved in conventional LC-MS analysis.

Accurate Mass and Time (AMT) tag proteomics is a high-throughput analysis methodology for LC-MS data that requires matching LC elution times and accurate masses of detected features to a database. To extend this approach to LC-IMS-TOF MS data, it was necessary to modify and create informatics tools to manage and process the enormous quantities of data generated, so the Smith team developed an expanded software package utilizing the AMT tag proteomics approach.

■ ■ ■

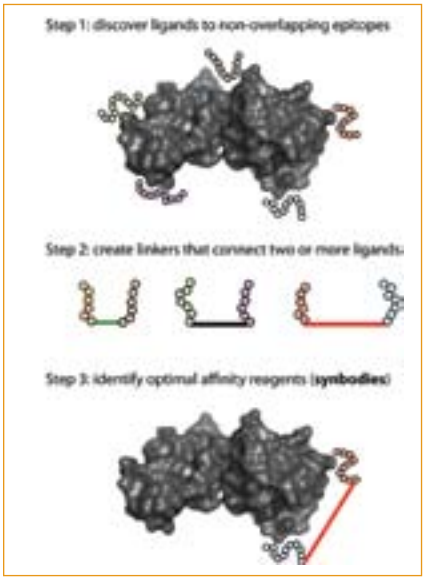
### Developing Synthetic Antibodies for Array-based Cancer Detection Principal Investigator

INVESTIGATOR:

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

Developing high-quality affinity reagents to the human proteome represents a grand challenge in basic and applied biomedicine. Reaching this goal will require transformative ideas that shift the current

**FIGURE 11. General strategy for the synthesis of synbodies**




paradigm away from methodologies that are costly and time consuming and that focus on novel solutions capable of revolutionizing the process by which protein capture reagents are created. In response to this challenge, the Chaput lab is working to develop a new class of synthetic antibodies that recognize protein targets with high affinity and specificity but do not require animal immunization or recursive selection steps for their discovery. Instead, their approach relies on a versatile two-step strategy in which noncompeting ligands are rapidly discovered and assembled into a series of multivalent protein capture reagents (Figure 11). 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. Dr. Chaput anticipates that these affinity reagents could be used in multiple formats to interrogate the nature of the human proteome and facilitate deeper insights into the molecular basis of human health and disease.

■ ■ ■

## Proteomic Characterization of Alternate Splicing and cSNP Protein Isoforms

INVESTIGATOR:

**Nathan Edwards**, Ph.D.
Georgetown University

Alternative splice and variant protein isoforms are often unobserved in proteomics datasets, due to fundamental limitations of the current proteomics informatics workflows. To address this issue, the Edwards team is developing informatics tools to improve the characterization of alternative splicing, coding single nucleotide polymorphisms (SNPs), as well as novel protein isoforms. The peptide sequence databases that the Edwards team developed have made it possible to capture evidence of alternative splicing, SNPs, and novel protein isoforms that would otherwise not be observed. A number of novel peptides and their associated proteins isoforms have been described because of this technique. The peptide sequence database used to find these novel peptides is updated quarterly, and it is freely available for download from the Edwards lab homepage: http://edwardslab.bmcb.georgetown.edu/.

The Edwards team has also developed a number of techniques that improve the number of confident peptide identifications that can be extracted from tandem mass spectra datasets. The HMMatch and PepArML tools demonstrate that it is possible to assign many more peptide identifications confidently, with no loss of statistical confidence, using multiple search engines and spectral matching. PepArML is currently the only open-source, freely available alternative to the increasingly popular commercial search engine combiner Scaffold. It is available as open-source software (http://peparml.sourceforge.net) and as a public service, using Edwards lab computer resources. 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. The public service is available at: https://edwardslab.bmcb.georgetown.edu/pymsio/.

The public Peptide Mapper Web-service (http://edwardslab.bmcb.georgetown.edu/ws/peptideMapper/) links peptides with their protein, transcript, and genomic evidence. This service makes it possible to evaluate the evidence for a novel protein isoform in the context of the available genomic, transcript, and protein evidence in external databases. The service also identifies the nucleotide sequence of peptides by alignment with transcripts and sends the nucleotide sequences to the University of California, Santa Cruz (UCSC) genome browser for alignment. The service can also, with one click, project a set of peptides into the UCSC genome browser for interactive evidence interpretation. This tool makes it possible to understand identified peptide sequences in the light of all of the protein sequence and genomic evidence for splicing and polymorphism.

■ ■ ■

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

INVESTIGATOR:

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

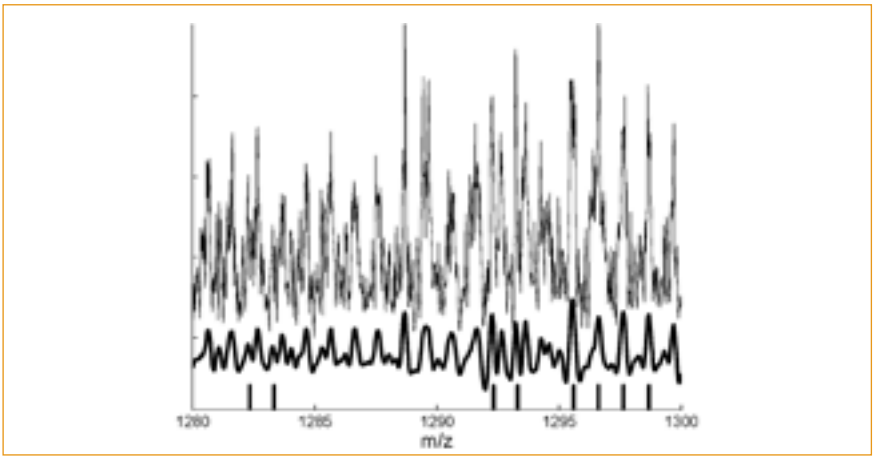
The Martin team has optimized a method to capture membrane proteins from adherent prostate cancer cells grown in culture. Optimization of the glycocapture conditions includes a variation of the oxidation and covalent biotinylation conditions to minimize cell toxicity. Additionally, the researchers have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

The team has also created a large database for over 500 glycoproteins and 2,000 peptides that were found in prostate cancer cell lines and xenograft tumors. This Prostate Cancer Database serves as the target source for future efforts to identify cancer-derived biomarkers in plasma.

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.

The Martin team has also developed the MaRiMba software tool to automate the creation of explicitly defined multiple reaction monitoring (MRM) transition lists required to program triple quadrupole mass spectrometers in such analyses. 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 12. The black tick marks indicate protein-induced signal in this high resolution MALDI TOF spectrum (light gray curve)**




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

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

Proteomic-assisted investigations of molecular and genetic function and discoveries of disease-related biomarkers rely on a growing list of technologically advanced assays. The Randolph team focuses its efforts 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, immunohistochemical labeling, and vibrational spectroscopies—also serve the goal of identifying proteins relevant to cancer processes and offer different strengths. Major projects by this team center on the analysis of LC-MS/MS, high-resolution MALDI-MS, and vibrational spectroscopies.

The Randolph team has developed an algorithm for detecting differentially expressed proteins between case and control groups. Based on semi-quantitative outputs from label free LC-MS/MS experiments, this algorithm has increased power across a large range of peptide abundance levels. The team has also developed a method to detect the statistically significant isotopic envelope of peaks from low-abundant peptides in high-resolution time-of-flight (TOF) spectra thereby enhancing the ability to detect peptide peaks in mass spectra.

To analyze data from vibrational spectroscopy platforms, the team has developed a statistical theory that clarifies how to exploit protein-induced signal. This theory allows statisticians to base their analysis more directly on proteomic information in noisy, high-dimensional spectroscopy curves (Figure 12). In a related project, the team has developed a statistical method to identify genomic and/or proteomic change-points in similar curve-like data.

### A Proteomics Approach to Ubiquitination

INVESTIGATOR:

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

Ubiquitin is a small molecule that regulates protein function by attaching to proteins in a process termed ubiquitination. Ubiquitination is one of the most common protein modification events in cells, and dysregulation of ubiquitination is involved in the pathogenesis of numerous types of cancers (Figure 13). However, there is no reliable method to globally analyze proteins that are modified by ubiquitin in cancer samples. The Peng team seeks to develop a proteomics approach to isolate and quantify ubiquitinated proteins from tumor tissues based on high-resolution MS. Once established, the method will be highly useful for profiling ubiquitinated proteins in mammalian samples, including clinical tumor tissues.

One challenge in analyzing ubiquitinated proteins is how to differentiate genuine ubiquitin-modified proteins from co-purified contaminants (i.e., false positives). The Peng team has developed a strategy to validate ubiquitinated proteomes, which allows the reconstruction of a "Virtual Western Blotting" image for every protein identified by mass spectrometry, thus detecting the signature of ubiquitinated proteins. The strategy provides a simple, effective method for quality control in analyzing ubiquitinated proteins. Moreover, the team has been developing a novel affinity matrix to capture ubiquitinated proteins from human samples with better efficiency.

> Clearly if you want to understand biology and you want to understand how the genome does what it does, you have to look in much greater detail at the proteins. The genome is after all the instruction book, but it doesn't do the work - the proteins do the work."
>
> Francis Collins, M.D., Ph.D.
> Director, National Institutes of Health

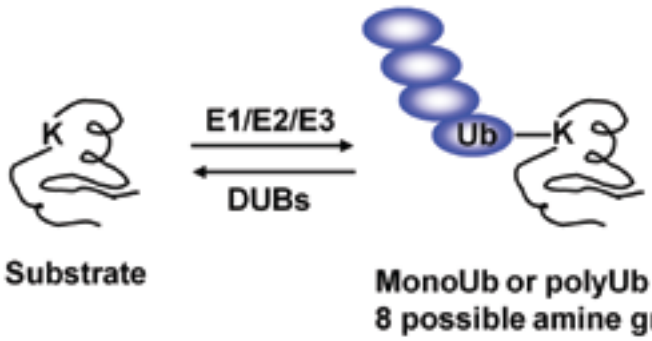
### Proteomic Phosphopeptide Chip Technology for Protein Profiling

INVESTIGATOR:

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

The Gao team is developing a proteomic phosphopeptide (PPEP) microchip technology platform that profiles proteins carrying phosphopeptide binding domains (PPBDs). 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. Current proteomic technologies identify proteins by either whole protein detection (2D electrophoresis and antibody binding) and/or peptide/phosphopeptide fragments (Mass Spectrometry). This team is developing tools that fill the gaps between these technologies by detecting/profiling phosphoprotein binding domain proteins (includes kinases, phosphatases and signaling proteins) systematically. These proteins are often active in cancer cells and are not detected by conventional proteomic methods easily.

**FIGURE 13. Protein ubiquitination plays a central role in cellular regulation and its dysfunction contributes to disease development**



The team has developed the µ Paraflo® Biochip Technology (Figure 14), which is an *in situ* synthesis technology that uses an activation agent generated using digital light and conventional chemistry on a microfluidic microchip. This unique µ Paraflo system differs from conventional DNA microarray technology in that it provides both a synthesis and an assay platform for peptides and proteins. Current efforts are screening fingerprints for phosphoprotein binding proteins such as SH2 domain containing proteins to develop clinical assays (diagnostic, prognostic, and/or monitoring drug treatment). The quantitative applications of the technology are applied to determining protein-binding constants in a massively parallel manner.

To lower the entry barrier of PepArray applications, the Gao team collaborated with the Li team (CPTC funding, collaborator at U. of Minnesota) to make a public/community Web tools source (google word: PepCyber) available. This resource includes PPEP for a database of human protein-protein interactions mediated by phosphoprotein binding domains and uPepArray Pro for PepArray design. The PepCyber PPEP (http://pepcyber.umn.edu/PPEP/) database focuses primarily on the human-curated interactions between binding domains in phosphoprotein binding proteins (PPBPs) and phosphopeptides (PPEP), and the uPepArray Pro helps users to design and organize peptides derived from PPEP and other Web sources into a microarray format for thousands of peptides to be made on chip for protein assays. The URL for this Web site is http://pepcyber.umn.edu/uPepArray/.

■ ■ ■

### Aptamer-Based Proteomic Analysis for Cancer Signatures

INVESTIGATOR:

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

The Walton team is developing an aptamer-based strategy for proteomics with better sensitivity and dynamic range than current array-based proteomic strategies. This aptamer-based strategy provides the option of ultimately using oligonucleotide microarrays to quantify proteomic signatures, which is advantageous for both infrastructure and reproducibility. With this cutting-edge technology, the team is developing a technique in which aptamers targeting different proteins are each labeled with a unique molecular barcode sequence (Figure 15), serving as unique identifiers of specific aptamers.

The team has 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.

■ ■ ■

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

INVESTIGATOR:

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

The Loo team seeks to develop a new technology platform that combines laser desorption and electrospray ionization, termed electrospray assisted laser desorption/ionization (ELDI), to provide top-down generated protein sequence tags to identify relevant cancer markers from biofluids and tissues. 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 are promising. The ability to combine on-line protein chemistries with

**FIGURE 14. µ Paraflo® Biochip Technology**

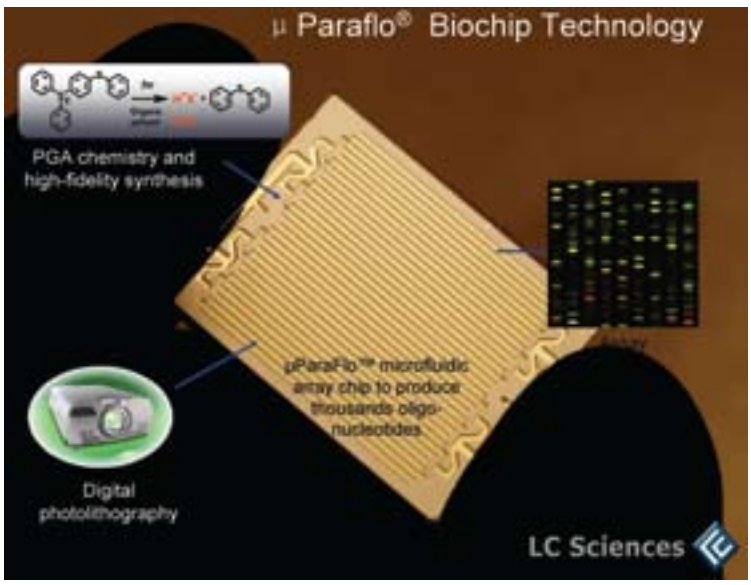
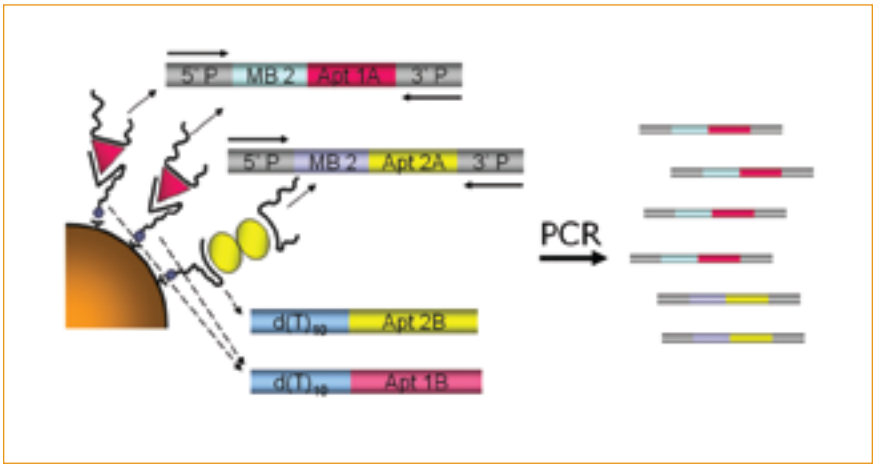


**FIGURE 15. Schematic of Aptamer-Based Proteomics Strategy**




separation technologies offers a unique platform for characterizing proteins, and it can potentially suggest new analytical advantages and strategies.

The team has developed protein pre-fractionation strategies to partition the salivary proteome for further characterization by top-down proteomics focusing on a protein-centric method (i.e., separation of intact proteins). Pre-fractionating by pI (isoelectric point) allows for larger protein loads without degrading gel electrophoresis resolution. This "divide and conquer" strategy has allowed this group to identify many more proteins than previously reported.

Extended MS/MS top-down experiments have been shown to be compatible with ELDI for proteins at high sensitivity (1 pmol). ELDI-MS shows a multiple charge distribution similar to that measured by conventional infusion ESI-MS. ELDI-MS sensitivity for intact protein analysis is slightly higher when compared to DESI and MALDESI.

■ ■ ■

### Global Production of Disease-Specific Monoclonal Antibodies

INVESTIGATOR:

Barry L. Karger, Ph.D.
Northeastern University

The Karger team has demonstrated the success of a method for generating and screening disease specific monoclonal antibodies (mAbs) to native glycoproteins in plasma. This method has been used to discover mAbs for biomarkers in lung cancer. A high-throughput platform to produce a large-scale Non-Small Cell Lung Cancer mAb library was established that led to the discovery of high-quality, well-characterized lung cancer specific mAbs. Unlike other methods that are based on recombinant proteins or peptide conjugates, this method can directly produce native form specific immuno-reagents for biomarker studies.

The team has also developed an antigen identification strategy with the goal of identifying the target antigens that correspond to these mAbs. To date, three of the antigens corresponding to mAbs have been identified (α chain haptoglobin, β 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.

■ ■ ■

### Analysis and Statistical Validation of Proteomic Datasets

INVESTIGATOR:

Alexey Nesvizhskii, Ph.D.
University of Michigan

The Nesvizhskii team is developing computational methods that take advantage of new mass spectrometry instrument capabilities, leading to increased accuracy and sensitivity. This increase in accuracy and sensitivity is particularly important in the identification of low abundance proteins where the confidence in the identification can be increased by combining multiple search tools and by integrating complementary spectral information from multistage MS.

The team has developed a new class of statistical models for improving peptide validation (sensitivity and robustness) 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).

The use of one data analysis search program can only identify a fraction of the spectra from an MS experiment. Variations in algorithms provide different identification results. To leverage this variation, a probabilistic framework tool was developed for combining results from multiple search engines. Using this tool, a significant gain in the number of peptides and proteins (more than 30 percent improvement) identified with high confidence was achieved. The

approach has been implemented in a computational tool called Scaffold (Proteome Software). The model will be available to the research community as part of the iProphet tool currently in development.

This team has also developed a novel statistical framework, called Q-Spec, to analyze label-free spectral count data. This 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. Q-Spec will be useful for designing label-free experiments that utilize spectral counting techniques.

■ ■ ■

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

INVESTIGATOR:

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

To construct and validate a software system for protein/peptide pattern-based biomarker discovery, the Mani team is combining peptide identity and pattern information obtained from high resolution and high mass accuracy spectra. The team is interested in enabling the robust quantification of unidentified peptide signals across many samples, leveraging peptide identifications via tandem MS to guide this process. The ultimate goal is to use this platform for discovering peptides or proteins that are differentially regulated in disease states.

The Mani team has developed a software tool called PEPPeR—A *P*latform for *E*xperimental *P*roteomic *P*atte*r*n *R*ecognition. PEPPeR uses high resolution and high mass accuracy LC-MS data and appropriately combines pattern-based (unidentified peptide peaks) and identity-based (peptides sequenced via MS/MS) information to generate peptide quantitation. PEPPeR, in its original form, was only capable of analyzing one-dimensional LC-MS/MS data. However, it is well known that two dimensional separations employing orthogonal separation modalities (e.g., reversed-phase separation of fractions produced by strong cation exchange 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. This software is freely available as a GenePattern module at http://www.broad.mit.edu/cancer/software/genepattern/desc/proteomics.html.

■ ■ ■

### PICquant – An Integrated Platform for Biomarker Discovery

INVESTIGATOR:

**Dennis Templeton**, Ph.D.
University of Virginia

The Templeton team seeks to develop an integrated software platform for protein quantification using a newly developed stable isotope mass tag to identify peptide peaks in treated urine samples accurately and, eventually, identify novel biomarkers for diseases. Urine is an easily accessed specimen that has not been well exploited for cancer biomarker studies.

The team has developed two new proteomic reagents. The first reagent is a stable isotope mass tag called PIC ($^{13}$C Phenylisocyanate) that has several advantages for protein quantification in complex mixtures. The second reagent, tentatively called TissueSolv, is capable of reversing many of the protein-protein crosslinks in formalin-fixed tissue. This reagent is superior to previous methods that mainly solublize the proteins, because it results in peptides whose masses match those predicted by protein databases. This approach enables the use of archival samples for validation of potential marker expression by MS.

A new software platform called PICquant has been developed that uses the PIC stable isotope mass tag for protein quantification. This platform includes: custom designed software (PICquant) to automatically quantify labeled peaks, a spectrum-comparison algorithm that groups spectra into a registry of spectra representing unique peptide families, and enhanced peptide sequencing capability. Completing the platform is a clinical registry that links acquired specimens to current and prospective clinical information including outcomes. 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. In addition to PICquant, a new software tool was developed called MAZIE. This software improved identification of peptide ion mass and charge.

## New Proteomic Algorithms to Identify Mutant or Modified Proteins

INVESTIGATOR:

David Tabb, Ph.D.
Vanderbilt University

Experiments to inventory complex collections of proteins often employ a "shotgun" strategy. Proteins are first digested to peptides by the trypsin enzyme, with resulting peptides separated through reversed phase liquid chromatography. Peptide ions are isolated in a tandem mass spectrometer and collided with gas to produce fragment ions that are recorded in MS/MS spectra. Typical experiments will produce 10,000 tandem mass spectra in an hour, necessitating algorithms to match these spectra to the peptide sequences that produced them.

The Tabb team produces tools to identify these MS/MS spectra, both by the conventional database search approach, and by the more flexible sequence tagging approach. Their intent is to improve the identification of peptides that differ by mutations from known sequences and by unusual chemical modifications. The group also focuses on the problems of filtering legitimate identifications from random ones and the problems of assembling protein lists from identified peptide lists.

Their novel "DirecTag" tool for sequence tag inference infers partial sequences from tandem mass spectra. This tool has been evaluated on a number of MS instruments ranging from TOF/TOFs to quadrupole ion traps. All results indicate that this algorithm could successfully interpret a larger fraction of tandem mass spectra than existing algorithms. DirecTag has been released with source code to the research community.

The real value of sequence tagging in cancer is the ability to identify mutations in cancer samples that have been missed by traditional database searches. The Tabb group has produced the "TagRecon" software to leverage DirecTag's output to identify mutations in peptides. The software has completed initial evaluation, and a publication should soon be available. The group has successfully discovered novel biology through use of TagRecon in the context of cancer cell lines and clinical proteomics data from colon cancer.

The team has developed a number of software tools currently in use throughout the CPTAC network. MyriMatch, a database search identification tool created by the Tabb Lab, was employed to identify the data produced through inter-laboratory studies of the CPTAC Unbiased Discovery Working Group. The scoring technique in MyriMatch has since been adapted for two other algorithms developed at Vanderbilt University and was incorporated in tools at two other institutions. MyriMatch makes more effective use of fragment ion intensities in comparison to other commercially available software tools.

Developed for protein assembly, IDPicker has proven to be invaluable in generating tables of spectral counts that can be used for identifying candidate biomarkers in large cancer datasets. The IDPicker tool enables users to organize experimental data into complex hierarchies. This protein assembler generated the identification reports underlying the papers produced by the CPTAC Unbiased Discovery Working Group.

## Enhancement of MS Signal Processing toward Improved Cancer Biomarker Discovery

INVESTIGATOR:

Dariya Malyarenko, Ph.D.
College of William and Mary

To increase the effectiveness of detection for protein/peptide molecules from matrix-assisted laser desorption-ionization (MALDI) time-of-flight (TOF) mass spectra (MS), the Malyarenko team is developing novel computational tools that can be used across laboratories employing this technology platform. The team's long-term goal is to deploy optimized data acquisition protocols and enhanced signal processing tools for the improvement of sample preparation and to streamline broad mass-range MS mining of proteomes and MS imaging for functionally important molecules related to different types of cancer.

The Malyarenko team has developed a number of open-source computational tools for the signal processing and statistical analysis of TOF mass spectra. New models have been created for signal and noise characteristics of this data. The team has also developed a procedure to record instrument parameters and experiment metadata systematically in order to minimize variability and bias in future experiments. Optimizing these acquisition parameters enhanced instrument sensitivity and experiment reproducibility.

Using error statistics analysis, the team combined several characteristics of mass spectra to optimize parameters for improved signal processing libraries. Applying these enhanced software libraries provided more than 10-fold increases in data compression, speed, and sensitivity of TOF MS signal detection. In addition, analytical framework and computational algorithms were developed for automated detection of charge states and multimers of molecular ions in the TOF domain (Figure 16). These tools enabled a fully automated self-calibration procedure, which achieved mass assignments with 10-fold accuracy over 5-fold broader mass-range in spectra of protein mixtures.

■ ■ ■

### Computational Tools for Cancer Proteomics

INVESTIGATOR:

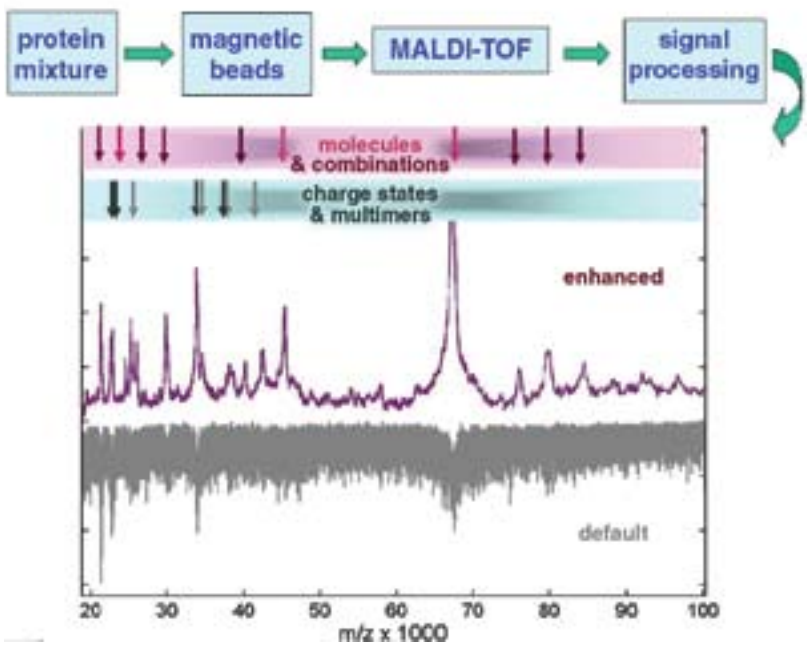
**William Old**, Ph.D.
University of Colorado at Boulder
*(Due to the passing of Dr. Kathryn Resing to cancer, Dr. William Old has assumed responsibility for this project.)*

Shotgun proteomics is a key technology in proteomic biomarker discovery and cancer signaling research that uses multidimensional peptide separation and gas phase peptide fragmentation to identify proteins in complex samples. With current technology, many thousands of proteins can be simultaneously identified and quantified in a single sample. However, serious limitations exist with respect to depth of sampling in complex mixtures, accuracy in assigning peptide sequences to MS/MS spectra, ambiguities in distinguishing protein isoforms, quantification of protein abundances, and characterization of posttranslational modifications such as phosphorylation. Dr. Old's team is developing new methods to profile protein expression and phosphorylation changes in response to signaling pathways and disease states, directly supporting studies of melanoma and prostate cancer.

In large-scale studies of complex samples, changes in phosphorylation are typically quantified by metabolic or chemical labeling of proteins in cells with isotopically distinguishable amino acids. Samples are mixed, and phosphopeptides are enriched on various affinity supports, followed by LC-MS/MS analysis. The Old team has developed a method for profiling phosphopeptides that does not depend on enrichment, and it could be performed quantitatively in complex samples in a label-free manner.

Dr. Old's team has also completed designing and implementing software for quantifying large-scale phosphoproteomics datasets that incorporate protein fractionation prior to reverse phase LC-MS/MS analysis. Their profiling method and software was applied to the detection of phosphorylation changes specific to metastatic melanoma, where they identified a phosphoprotein of unknown function as necessary for metastatic growth of melanoma cells (Old, W.M. *et al.*, 2009 *Molecular Cell* 34,115-131).

**FIGURE 16. Computational algorithms are able to separate peaks from noise thereby enhancing MALDI-TOF spectrum**



# Proteomic 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 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, the 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 Proteomic 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. The Reagents Data Portal is available at http://antibodies.cancer.gov.

■ ■ ■

### Reagents

The Reagents Data Portal is in the process of expansion as the program makes way for a great number reagents in the pipeline that are needed for effective proteomic analysis. To date, more than 25 antigens and 75 monoclonal antibodies have been generated against human cancer-associated proteins and each antibody is added to the web portal along with corresponding characterization data (isotype, western blot, ELISA, immuno-mass spectrometry, surface plasmon resonance, immunohistochemistry, and nucleic acid programmable protein arrays).

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

The NCI Clinical Proteomic Measurement Assessment Materials Program at NIST is developing and documenting the characterization of complex biological mixtures for the evaluation of proteomic analysis platforms by the CPTAC Network. The goal is to develop proteomic standard reference materials (SRM) and provide them to the scientific community through the CPTC Reagents and Resources program.

NIST has been preparing and characterizing reference material, including yeast, plasma, and cell lysates, which may be spiked with exogenous proteins or peptides. These reference materials serve as important reagents for the program and support proteomic analysis by mass spectrometry and other analytical techniques such as affinity-based technologies. Such high quality, well characterized, and readily available 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.

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 1). The Yeast Lysate was used in several CPTAC inter-laboratory studies. The Peptide Performance Mixture will be an aqueous mixture of approximately 400 synthetic peptides, whose concentration will span approximately three orders of magnitude. The generation of SRMs requires stability testing, quality control evaluation, and extensive characterization before being released to the scientific community.

**TABLE 1. New Products Approved for Reference Material Development**

| Standard Reference Material | Purpose |
|---|---|
| Yeast Lysate for Proteomic Research (SRM 3953) | Model proteome of moderate complexity intended for evaluating the measurement quality of proteomic investigations |
| Peptide Performance Mixture for MS (SRM 3592) | Complex mixture of peptides intended for evaluating the performance of MS instruments performing quantitative data-dependent acquisition |

### *Antigen Production*

A key component of the CPTC program is the development of customized reference protein standards for MS assessment studies. Argonne National Laboratory (ANL) is producing, qualifying, and characterizing proteins for CPTC. Recombinant protein production at ANL serves a dual purpose for the scientific community: All ANL recombinant proteins are 15N-labeled for use in MS studies within the CPTC program, while simultaneously used as antigens to generate monoclonal antibodies.

The DNA clones that are used to create the recombinant proteins at ANL are currently deposited at Arizona State University's DNASU Plasmid Repository and formally deposited at Harvard University in the Plasmid Information Database (PlasmID) and are available to the scientific community. A database is maintained that contains detailed annotation for each plasmid that can be accessed by users through the Web site. The goal is to collect plasmids, fully sequence validate them, and then distribute them worldwide. Additionally, proteins are now being made available for purchase at ANL (antigens.anl.gov).

### *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 Polanski and Anderson in 2006 (Polanski M, Anderson N.L. A list of candidate cancer biomarkers for targeted proteomics. *Biomarker Insights.*, 2007;1:1-48.) as part of collaboration between CPTC and several laboratories and companies. Antigens produced by the ANL are provided to institutions contracted for antibody generation through requests for proposals (RFP) mechanism. For each RFP, the contractor receives 40 endotoxin free purified proteins from ANL and in turn generates 10 monoclonal IgG antibody supernatants for each target protein. After initial characterization studies, three of the 10 antibodies are selected to be grown up and purified. The purified antibodies are evaluated for in-depth characterization at four collaborating centers: NCI-Frederick, NCI's Center for Cancer Research Tissue Array Research Program (Gaithersburg, MD.), the Arizona State University (Tempe, AZ), and the Human Protein Atlas at KTH-Royal Institute of Technology (Stockholm, Sweden) (Figure 17).

All of the hybridomas are deposited at the Developmental Studies Hybridoma Bank (DSHB) at the University of Iowa where they are made available to the public at nominal cost. A key advantage of these antibodies is the diversity of characterization data available.

**FIGURE 17. CPTC Antibody Characterization Pipeline**

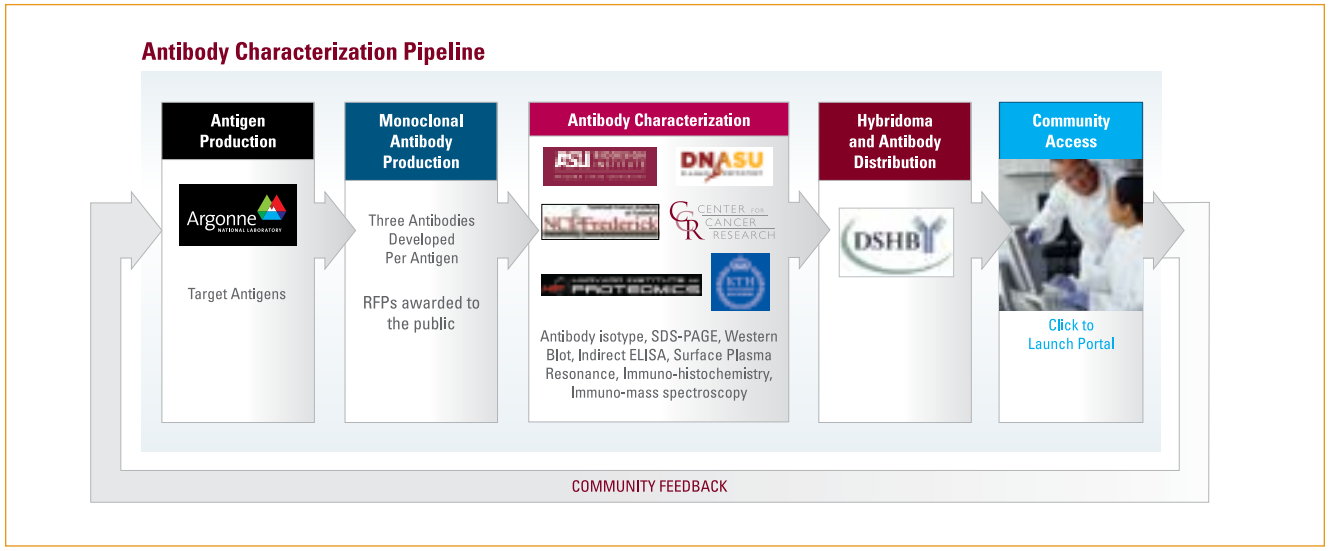
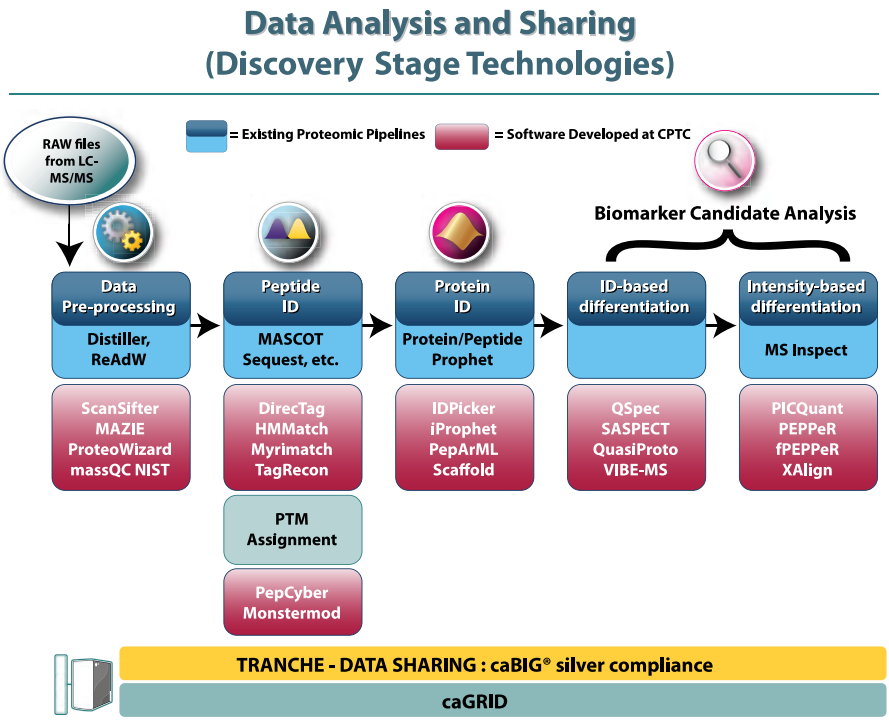


**FIGURE 18. CPTC Developed Software Tools Used Throughout the Biomarker Discovery Stage Pipeline**




## Resources

### *Storage and File Sharing – Tranche*

Tranche, an open-source file sharing tool for scientific data, is the primary repository for proteomics data (proteomecommons.org/tranche). Tranche allows collaborators to share research data regardless of file size, the number of files or file format. Tranche can also securely share data in pre-review before publication. Files can be encrypted and only made available to the original uploader and with whomever they share it. In this way, researchers can share data without worrying about compromising data security. 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 received caBIG® Silver Level Certification, making CPTC data accessible to the broader biomedical research community.

### *Standard Operating Procedures (SOPs)*

Working closely with the NCI Office of Biorepositories and Biospecimen Research (OBBR), CPTC developed a biospecimen collection protocol, which is the first multicenter SOP generated by the program. This SOP resulted from the consolidation of protocols from five leading clinics and associated collaborators.

### *Software Packages*

One of the major outputs from CPTC has been the development of software analysis tools. Analyzing MS data for protein identification includes a number of steps (Figures 18 and 19). Briefly, the raw mass spectra are first processed to improve the quality of the spectra. Poor spectra are discarded. Next, peptides are identified from the spectra. If one is searching for post-translational modifications, those would also be identified at this point. After that, protein identities are inferred from the identified peptides. Finally, a number of quantitative and semi-quantitative methods are available to differentiate proteins upregulated in specific disease states. These disease-linked proteins may then comprise a biomarker candidate list. Throughout this pipeline, MS data can be stored in the ProteomeCommons.org Tranche network for file sharing among collaborators and secure storage. See below for further descriptions of these tools.

### Data Pre-processing Software

SCANSIFTER: The "ScanSifter" algorithm, a Vanderbilt-developed software, assesses the quality of each raw spectrum and discards poor quality spectra. As such, this application streamlines data analysis systems

MAZIE (MASS AND CHARGE (Z) INTERFACE ENGINE): This software improves identification of peptide ion mass and charge, based on the isotopic distribution of peptide ion envelopes. It will be distributed freely to the research community upon publication.

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 liquid chromatography–mass spectrometry (LC-MS) dataset computations.

MASSQC: MassQC is a software package that serves to diagnose mass spectrometry 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. 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 Software

DIRECTAG: DirecTag identifies peptides through sequence tagging using automated sequence tag inference that has been shown to be an accurate way to identify peptide sequences from tandem mass spectra. DirecTag has been released with source code to the research community.

HMMATCH: The 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.

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 datasets of the CPTAC Unbiased Working group.

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 Software

PEPCYBER: The PepCyber database focuses primarily on the interactions between binding domains in phosphoprotein binding proteins (PPBPs) and phosphopeptides (PPEP).

MONSTERMOD: MonsterMod matches a user-supplied list of peptide or protein sequences to a collection of tandem mass spectra.

### Protein ID Software

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 datasets. It has also been instrumental in organizing the complex datasets from the CPTC 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 allows the combining of results from multiple search tools and also takes into account other supporting factors to compute a new probability.

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.

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.

■ ■ ■

### ID-based Differentiation Software

QSPEC: QSpec is for data generated by the spectral count method that has become an accepted method for label-free quantitation in proteomics.

SASPECT: SASPECT provides a function for identifying differentially expressed proteins between two sample groups using spectral counts from liquid chromatography–tandem mass spectrometry (LC-MS/MS) experiments.

QUASIPROTO: QuasiProto is designed for spectral count differentiation in complex proteomic datasets and reports the numbers of spectra matched to each protein.

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 mass spectrometry 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.

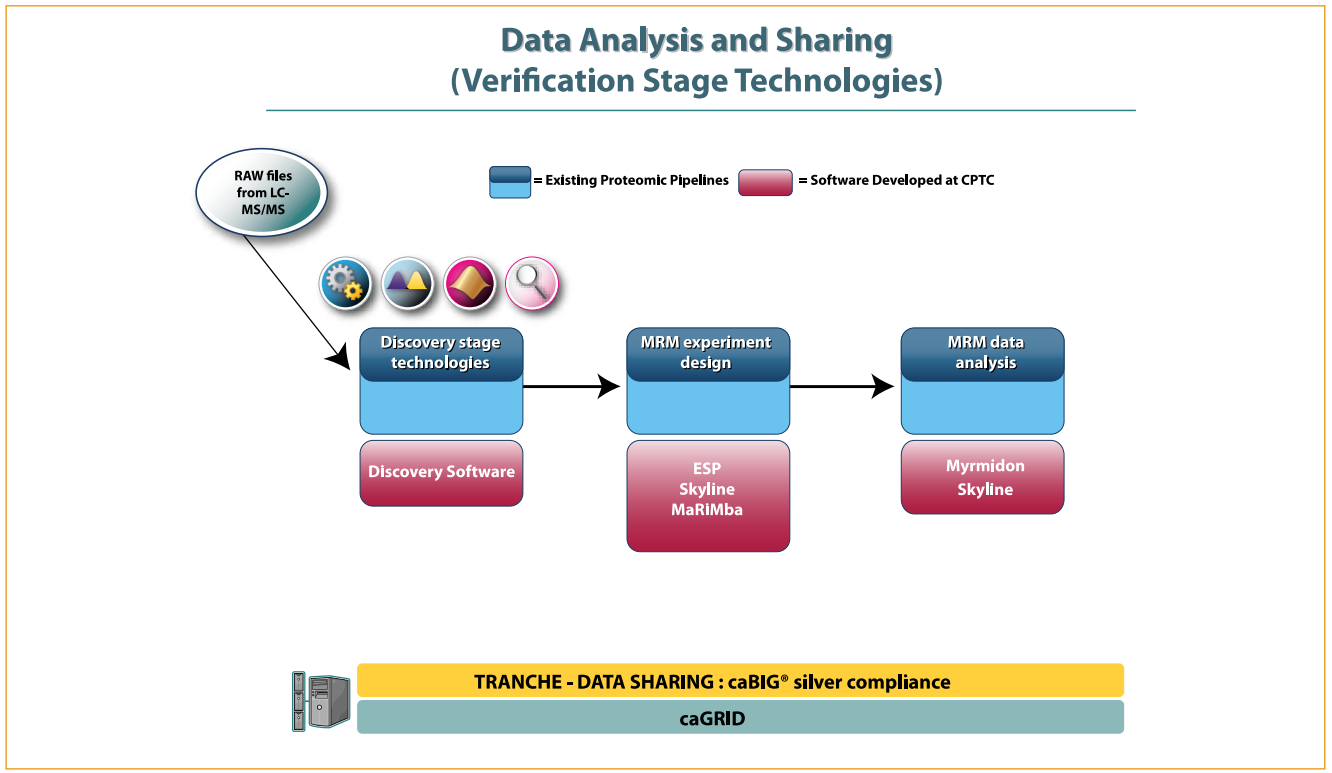
■ ■ ■

### Intensity-Based Differentiation Software

PICQUANT: PICquant is an integrated software platform for biomarker discovery using a new stable isotope mass tag ($^{13C}$ Phenylisocyanate, PIC). The platform has a clinical registry that links acquired specimens to current and prospective clinical information. 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.

PEPPER: A Platform for Experimental Proteomic Pattern Recognition, (PEPPeR) has been developed that uses high resolution and high mass accuracy LC-MS data and combines pattern-based (unidentified peptide peaks) and identity-based (peptides sequenced via MS/MS) information for peptide quantitation.

**FIGURE 19. CPTC Developed Software Tools Used Throughout the Biomarker Verification Stage Pipeline**



FPEPPER: This is a recently-developed extension of PEPPeR that 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. This software is freely available as a GenePattern module.

XALIGN: XAlign is a two-step alignment algorithm that detects significant peaks that are common to all samples and aligns them using refined m/z and retention time variation values.

While many protein biomarker candidates have been discovered, few have translated into biomarkers of clinical utility. CPTC has developed a time-efficient, cost-effective method of triaging candidate lists of proteins into well-credentialed, high-priority biomarker candidates. A major effort of CPTC has been to create the software

## CPTC 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 datasets. 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 biomedicine 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 datasets. 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 mass spectrometry raw files, Tranche is the primary 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, 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 machine. The pipeline integrates with leading search engines. 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.

for this so-called "verification" method. Briefly, verification assays are built for proteins received from discovery proteomics technologies. Then, a multiple-reaction monitoring (MRM) experiment is conducted. A number of CPTC software tools now exist to ease both the experimental setup of this experiment, as well as the analysis of the data. See below for further descriptions of these tools.

### MRM Experiment Design

ESP: The first step in developing effective MRM assays is to identify the appropriate peptides. Identifying 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 for use in targeted MS–based assays.

SKYLINE: Skyline is a Windows client application for building MRM methods for large-scale proteomics studies and analyzing the resulting mass spectrometer data.

MARIMBA: The MaRiMba software tool automates the creation of explicitly defined MRM transition lists for native and isotopically heavy peptides which are required to program triple quadrupole mass spectrometers in such analyses. This open-source application is operated through a graphical user interface incorporated into the TransProteomic 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 MRM experiments. Scaling across multiple replicates in multiple cohorts, the software assists in recognizing and integrating chromatographic peaks, producing reports suitable for statistical interpretation.

# Small Business Innovation Research (SBIR)

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

To maximize the program's capabilities and impact, CPTC partners with the biotechnology industry via NCI's Small Business Innovation Research (SBIR) Program, a contract mechanism that supports early stage research and development by small businesses. Through this 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.

## ACCACIA INTERNATIONAL, INC.

### High-Throughput Selection of Aptamers Against Cancer Biomarkers

Aptamers (nucleic acid or peptide molecules that bind to a specific target) 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 International, Inc. is generating aptamer receptors and developing aptamer-based diagnostic assays such as the proximity ligation assay (PLA) that have much greater sensitivity than conventional ELISAs. The company has selected aptamers against three targets with low nanomolar Kds: EGFR, ErbB2, and FGFR1.

## ALLELE BIOTECHNOLOGY & PHARMACEUTICALS, INC.

### Yeast Single Chain Antibodies as Capture Reagents

Allele Biotechnology & Pharmaceuticals, Inc. is developing 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. Therefore, scFV can serve as alternative capture reagents with high specificity, high affinity, and small size and 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 complimentarity-determining region sequences with designed codons mimicking natural human antibody diversity. The isolated antibodies will be characterized by Western Blot, ELISA, and microarray platforms.

The company has isolated multiple high affinity and specificity antibodies against eight of the 10 proposed antigens, and pools of antigen-specific scFv clones against the other two.

### Expression of Mammalian Glycoproteins Using modified BEVS

The objective of this proposal is to develop and validate novel systems for efficient expression of cancer-related human glycosylated proteins. The availability of such systems will aid analysis of cancer related proteins, especially those of low abundance from bodily fluids, and accelerate the development of effective cancer diagnostics and therapeutics.

The baculovirus expression vector systems (BEVS) using insect cells provide an alternative at lower cost and faster turnaround time to mammalian systems. The company will develop a complete system to utilize BEVS in both insect and mammalian cells with unique purification methods and flexibility to produce glycosylated proteins.

## INTEGRAL MOLECULAR, INC.

### Mapping of Epitopes on Cancer Biomarkers

Monoclonal antibodies (mAbs) are important therapeutic, diagnostic, and research reagents in the field of oncology. Characterizing their binding sites on target antigens can elucidate cancer-specific topological arrangements, and can aid in the development and selection of optimized mAbs. However, there are currently no commercial tools available that enable automated, rapid functional analysis of mAb epitopes in membrane proteins and other structurally-complex cancer biomarkers. Integral Molecular is developing a novel technology that addresses the bottleneck in conventional mutational analyses: the expression and analysis of large libraries of point-mutated proteins. In this proposal, the company will use this technology to map the epitopes of mAbs directed against structurally-complex cancer biomarker candidates that are resistant to direct structural analysis.

## INTRINSIC BIOPROBES, INC.

### Multiplex Mass Spectrometric Immunoassays

Intrinsic Bioprobes, Inc. is developing and validating multiplex mass spectrometric immunoassays (MSIA) for detecting and qualifying cancer-related proteins with intrinsically low concentrations in bodily fluids. The company has created a multiplexed MSIA that is able to detect and profile five plasma proteins simultaneously, with a concentration range of 1 ng/mL to 1 mg/mL. The company is preparing to start phase II of the project, the objective of which is to develop a fully functional quantitative, automated, high-throughput, multiplex affinity protein capture technology platform and Multiplex Mass Spectrometric Immunoassays for the analysis of low abundance cancer related proteins/peptides from bodily fluids.

### High-Throughput Mass Spectrometric Epitope Mapping

For this project, the company is proposing a high-throughput mass spectrometry-based analysis of epitope determinants to identify the interacting protein sequences using immunoaffinity capture combined with enzymatic digestion of the target protein; either pre- or post-capture. The project will develop reproducible methods to demonstrate epitope mapping for monoclonal antibodies of interest to NCI tested by multiple sampling and quantifiable results. Mapped epitopes will be functionally characterized by amino acid substitutions in the epitope sequence using high-throughput spotted array surface plasmon resonance imaging and SPR-Biomolecular Interaction Analysis to identify the key amino acids that are critically involved in binding the antibody to the antigenic epitope.

## LIFESENSORS, INC.

### Novel Protein Expression Technologies for Glycoproteins

This project will use SUMO-fusion vector to improve protein expression and protein secretion and engineered humanized *Pichia pastoris* strain to control post-translational glycosylation. First they will demonstrate that i) this novel system is superior in increasing protein expression and secretion, facilitating protein purification and generating desired N-terminal amino acid, ii) produce proteins with more than 90% homogeneity, iii) produce proteins with mammalian-like N-glycan complex structures. Once their system is validated by fulfilling above criteria, they will express and purify glycosylated proteins.

## MESO SCALE DIAGNOSTICS, LLC.

### Automated Multi-Array Platform for Cancer Biomarkers

Meso Scale Diagnostics, LLC is developing a quantitative, automated, high-throughput, multiplexed immunoassay platform, which will simultaneously measure up to 75 biomarkers using 30 µl 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 60–300 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. The company is preparing to start phase II of the project.

## PREDICTIVE PHYSIOLOGY AND MEDICINE, INC. (PPM)

### Immunoaffinity Capture Coupled with Ion Mobility Spectrometry

Predictive Physiology and Medicine, Inc. (PPM) will produce an efficacious biomarker panel for prostate cancer and develop instrumentation that can be used directly in the clinical laboratory.

## QUADRASPEC, INC.

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

Quadraspec, Inc. is developing an array of 10–100 microfabricated disposable nano-immmunoaffinity chromatography (NIAC) columns originating from a central 10–50 µl well that can be used to selectively capture and enrich 10–100 cancer markers from bodily fluids as they flow into the column array from the central well where the sample is deposited. Quadraspec predicted that as a sample flows into the NIAC columns, this loading process will enrich and extend the dynamic range as well as enhance sensitivity by a million-fold while minimizing non-specific binding. Detection is performed using conventional monitoring methods such as conjugated secondary antibodies and enzyme amplification. The goal is to develop a simple chip that can be used in clinical labs.

## RANA BIOSCIENCES

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

Protein phosphorylation is a central mechanism of cellular regulation and aberrant phosphorylation activities are involved in the development of human cancers. Highly specific capturing reagents are needed to detect and monitor such 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-specific phosphorylated proteins. In phase I, the objective is to develop an *E.coli* cell-free system that will produce proteins containing phosphotyrosines.

### An Expression System for Synthesis of Glycoprotein with Defined O-Glycan Structure

The goal of this project is to establish a cell-free translation system that can simply and economically produce glycoproteins with defined O-glycan structures. O-glycans will be incorporated into proteins through translational suppression at desired site(s). The main objective of the project is to develop amino acyl-tRNA synthetases (aaRS) that are able to activate glycoamino acids carrying structurally defined O-glycans.

## RULES-BASED MEDICINE, INC. (RBM)

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

In phase I, Rules-Based Medicine, Inc. (RBM) developed a quantitative, automated, 5-plex immunoassay for the rapid detection of low abundance cancer-related proteins. In phase II, RBM will validate the immunoassay by screening an additional 45 targets. The final product will result in a 50-plex immunoassay for low abundance cancer-related proteins. RBM is currently in its phase II and has completed most of the antibody screening for the project. They have developed a number of assays and are planning to scale up and commercialize them.

## SEQUENOM, INC.

### Sensitive Protein Detection Combining Mass Spectrometry and PCR

Sequenom, Inc. proposed a novel approach to detect and quantify proteins in bodily fluids that integrates three technologies in which the company and their collaborators have expertise: immuno-PCR, competitive-PCR and mass spectrometric analysis using Sequenom's MassARRAY platform.

The company has conducted proof of concept experiments and compared their 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.

# OUTREACH ACTIVITIES

## Meetings and Workshops

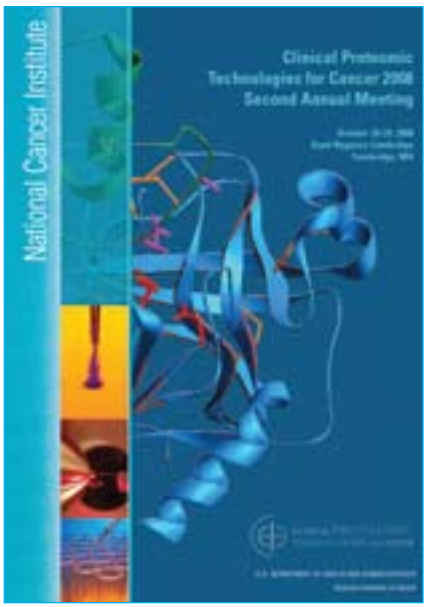
### Second Annual Meeting

CPTC held its second annual meeting in Cambridge, MA on October 28–29, 2008, bringing together more than 200 participants representing the full gamut of scientific fields that contribute to the program'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 program components.

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 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. 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.

In his closing remarks, CPTC Director Henry Rodriguez, Ph.D., M.B.A., noted that the program 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 program, it showed 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.




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

Working with the international community, CPTC has led efforts to provide recommendations for rapid proteomics data release and sharing policies that are similar to the Bermuda Principles, which served as a catalyst in the world of genomics. The lack of such policies in proteomics is currently seen by many as a stumbling block for the progress and support of the field as a whole.

In August 2008, CPTC sponsored a summit in The Netherlands for members of the international proteomics community to define what it would take to have proteomics data released into the public domain as soon as they are produced.

***What type of proteomic data should be released? What types of metrics define data quality for proteomics? When should data be submitted?*** These are difficult questions, given that this is a complex and burgeoning field, but they must be addressed now if proteomics is to ever live up to its promise. Taking part in these discussions were representatives from European Union funding agencies, the National Cancer Institute/National Institutes of Health, European Bioinformatics Institute, The Wellcome Trust, Genome Canada, National Center for Biotechnology Information, National Institute of Standards and Technology, proteomics journals, and many international universities.

***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 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.

***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.

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. Central repositories, for their part, should clearly define minimum submission requirements, encourage rich annotation, and develop seamless submission procedures.

This international one-day summit was a major step forward for the proteomics community. It is anticipated that the principles developed at this meeting can be readily adopted by the field as guidelines for releasing and sharing proteomics data.

**To learn more about how CPTC is leading data sharing efforts in the proteomics research community, please visit the Data Sharing page on the CPTC Web site at http://proteomics.cancer.gov.**

Rodriguez, H., *et al.* Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: The Amsterdam Principles. *J. Proteome Res.*, 2009, 8 (7), p 3689.

Eastman, Q. Proteomics researchers solidifying principles for data sharing. *J. Proteome Res.*, 2009, 8 (7), p 3220.


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

An NCI-FDA Interagency Oncology Task Force (IOTF) was established in 2003, with a Molecular Diagnostics subcommittee formed in 2007. A workshop by this subcommittee was held on October 30, 2008 in Cambridge, MA. The purpose of this workshop was to discuss requirements for analytical validation of protein-based multiplex ("bridge") technologies in the context of their potential clinical use.

This workshop focused on technology-specific analytical validation processes to be addressed prior to use in clinical settings. The workshop used a case study approach that discussed issues related to:

- Validating a proteomic technology
- Specimen and population issues
- Statistical issues
- Understanding the regulatory pathway to commercialization

A summary document has been submitted for publication, which will discuss analytical validation issues that specific proteomic technologies should address when seeking FDA approval.

## FDA Documents – "Mock 510(k)"

Key outcomes from the NCI-FDA IOTF Molecular Diagnostic workshop were the need to create documents that would help orient the FDA and scientific community to protein-based multiplex assays in novel diagnostics. As a result, mock 510(k) documents (non-regulatory documents in the form of 510(k) submissions) were drafted and are being submitted for publication, one for protein-based multiplex MS platforms and the other for protein-based multiplex affinity platforms. These mock 510(k) documents will help orient the FDA to protein-based multiplex assays in novel diagnostics and serve as a springboard for guidance to the proteomics community.

> "There's really no guidance for multiplex proteomic assays...There are unique issues when you start to do a multiple test in a single tube or platform."
>
> Elizabeth Mansfield, Ph.D.
> Senior Policy Analyst
> Office of *In Vitro* Diagnostic Devices, FDA

## *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.


## Patient Corner

CPTC recognizes the important role that patients, patient advocates, and other members outside of the traditional science community play in advancing cancer research, and is actively involved in dialogue with such communities. A number of educational resources have been made available under a new Patient Corner section of the CPTC Web site. The inextricable link between proteomics research and patient outcomes is highlighted through podcasts, webinars, brochures, and tutorials. This site can be accessed at: http://proteomics.cancer.gov/library/patientcorner

## Advocacy Outreach

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.

> "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."
>
> Elda Railey
> Co-Founder
> Research Advocacy Network

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

OAR | Office of Advocacy Relations

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

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.

*Elizabeth Nielson*
*CARRA Program Manager*
*Office of Advocacy Relations*
*National Cancer Institute*

## Under-Represented Students

The NIH 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) 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:

- Create a pipeline of underserved students and investigators in the fields of emerging and advanced technologies
- 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
- Enhance the application of emerging technologies to cancer research through increased training and educational opportunities
- Foster academic, scientific and multi-disciplinary research excellence to culminate the emergence of a mature investigator capable of securing competitive advanced research funding

**CRCHD**
**Center to Reduce Cancer Health Disparities**

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 and a member of the CPTAC Network, is an active participant in ET CURE program. She is currently giving three budding students—two undergraduates and one high school student—exposure to the field of clinical proteomics with the goal of getting them excited about the field. The curriculum for these students includes a number of activities geared towards preparing them for a successful career in biomedical research.

*Tao Large, Tim Nguyen, and Brianna Byers are currently students in the ET CURE program at FHCRC.*

# PARTNERSHIPS

## NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST)

NIST is developing mass spectrometry assessment and proteomic standard reference materials through an Interagency Agreement with CPTC.

## TEXAS A&M UNIVERSITY NATIONAL INSTITUTE OF STATISTICAL SCIENCES (NISS)

CPTC works with Texas A&M and NISS to support expertise in experimental study design, metrology, statistical analysis, and methodological approaches to applying proteomic technology platforms toward clinical measurement.

## ARGONNE NATIONAL LABORATORIES (ANL)

CPTC has an Interagency Agreement with ANL to produce 15N labeled cancer-related proteins for use in antibody production, affinity capture technology development, and creation of protein standards.

## HUMAN PROTEIN ATLAS (HPA)

CPTC has partnered with HPA to further characterize monoclonal antibodies generated by CPTC using tissue microarrays.

## HARVARD INSTITUTE OF PROTEOMICS (HIP) DNASU PLASMID REPOSITORY OF THE BIODESIGN INSTITUTE AT ARIZONA STATE UNIVERSITY (DNASU)

HIP and DNASU serve as central repositories for plasmid clone collections and distribution for CPTC.

## VIRGINIA G. PIPER CENTER FOR PERSONALIZED DIAGNOSTICS OF THE BIODESIGN INSTITUTE AT ARIZONA STATE UNIVERSITY

The Virginia G. Piper Center for Personalized Diagnostics of the Biodesign Institute at Arizona State University applies Nucleic Acid Programmable Protein Arrays (NAPPA) technology to monoclonal antibodies from the CPTC program.

## DEVELOPMENTAL STUDIES HYBRIDOMA BANK (DSHB) AT THE UNIVERSITY OF IOWA

The DSHB, created by the NIH as a national resource, collects, stores, grows, and distributes all hybridomas and monoclonal antibodies generated by CPTC.

## FOOD AND DRUG ADMINISTRATION (FDA)

The Food 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 was established between CPTC (NCI) and FDA, which focuses on accelerating proteomic technology development and application in clinical settings.

## AMERICAN ASSOCIATION OF CLINICAL CHEMISTRY

CPTC has entered into a memorandum of understanding with the AACC to join forces on promoting and educating the clinical chemistry community in the area of proteomic standards and technology advances.

## KOREAN INSTITUTE OF SCIENCE AND TECHNOLOGY (KIST)

CPTC has entered into a memorandum of understanding with KIST to promote proteomic technology optimization and standards implementation in large-scale international programs.

## IMAGENES AND MILLIPORE

imaGenes and Millipore have chosen to distribute select monoclonal antibodies created and characterized by CPTC.

## SMALL BUSINESS INNOVATION RESEARCH (SBIR)

CPTC has partnered with the small business community via the NCI's SBIR program, a contract mechanism that supports early stage research and development by small businesses. Through the SBIR program, CPTC encourages and enables companies developing proteomic technologies and platforms to develop standardized, well-characterized reagents in the commercialization of new tools and kits for the cancer community.

## INNOVATIVE MOLECULAR ANALYSIS TECHNOLOGIES (IMAT)

CPTC has partnered with NCI's IMAT program to further promote integrated technology development in cancer proteomics research.

# APPENDIX

## Organizations Participating in the CPTC Initiative

Accacia International, Inc.
Allele Biotechnology & Pharmaceuticals
American Association of Clinical Chemistry
Argonne National Laboratory
Battelle Pacific Northwest Laboratories
Buck Institute for Age Research
California Pacific Medical Center
The College of William and Mary
Developmental Studies Hybridoma Bank at the University of Iowa
Discovery Park at Purdue University
Emory University
Epitome Biosystems, Inc.
Epitome, Inc.
Epitomics Inc.
European Bioinformatics Institute
Food and Drug Administration
Fred Hutchinson Cancer Research Center and its clinical and research partners, the University of Washington and Children's Hospital and Regional Medical Center
Georgetown University
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)
Imagenes
Indiana University
Indiana University–Purdue University Indianapolis
Indiana University School of Medicine
Institute for Systems Biology
Integral Molecular, Inc.
Intrinsic Bioprobes, Inc.
Korean Institute of Science and Technology
Lawrence Berkeley National Laboratory
LifeSensors, Inc.
Massachusetts Institute of Technology
Memorial Sloan-Kettering Cancer Center
Meso Scale Diagnostics, LLC
Michigan State University
Millipore
Monarch Life Sciences
National Cancer Institute–Center for Cancer Research Tissue Array Program
National Cancer Institute–Frederick Advanced Technology Program
National Cancer Institute–Innovative Molecular Analysis Technologies
National Institute of Standards and Technology
National Institute of Statistical Sciences
New York University Langone Medical Center
Northeastern University
Pacific Northwest Laboratories
The Plasma Proteome Institute
Predictive Physiology and Medicine, Inc. (PPM)
Purdue University
Quadraspec, Inc.
Rana Biosciences
Rules-Based Medicine, Inc. (RBM)
Sequenom, Inc.
Texas A&M University
The Broad Institute of MIT and Harvard, Proteomic Platform and Cancer Program
The University of Texas M.D. Anderson Cancer Center
University of British Columbia
University of California, Los Angeles
University of Colorado at Boulder
University of Houston
University of Michigan
University of Victoria
University of Virginia
University of Washington
Vanderbilt University
Virginia G. Piper Center for Personalized Diagnostics of the Biodesign Institute at Arizona State University

## CPTC Publications


Addona, T. A., *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol.* 2009; 27(7): 633-41.

Ahn, N. G., *et al.* Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem Biol.* 2007; 2(1): 39-52.

Alves, P., et al. Fast and accurate identification of semi-tryptic peptides in shotgun proteomics. *Bioinformatics.* 2008; 24(1): 102-9.

Anderson, N. L., *et al.* A human proteome detection and quantitation project. *Mol Cell Proteomics.* 2009; 8(5): 883-6.

Anderson, N. L., *et al.* SISCAPA peptide enrichment on magnetic beads using an in-line bead trap device. *Mol Cell Proteomics.* 2009; 8(5): 995-1005.

Anderson, N. L. The Clinical Plasma Proteome: A Survey of Clinical Assays for Proteins in Plasma and Serum. *Clin Chem.* 2009.

Arnett, D. R., *et al.* A proteomics analysis of yeast Mot1p protein-protein associations: insights into mechanism. *Mol Cell Proteomics.* 2008; 7(11): 2090-106.

Baker, E. S., *et al.* Simultaneous fragmentation of multiple ions using IMS drift time dependent collision energies. *J Am Soc Mass Spectrom.* 2008; 19(3): 411-9.

Belov, M. E., *et al.* Dynamically multiplexed ion mobility time-of-flight mass spectrometry. *Anal Chem.* 2008; 80(15): 5873-83.

Burgess, E. F., *et al.* Prostate cancer serum biomarker discovery through proteomic analysis of alpha-2 macroglobulin protein complexes. *Proteomics Clin. Appl.* 2008; 2: 1223-1233.

Cao, Z., *et al.* Use of fluorescence-activated vesicle sorting for isolation of Naked2-associated, basolaterally targeted exocytic vesicles for proteomics analysis. *Mol Cell Proteomics.* 2008; 7(9): 1651-67.

Cao, X., *et al.* Improved sequence tag generation method for peptide identification in tandem mass spectrometry. *J Proteome Res.* 2008; 7(10): 4422-34.

Carr, S. A., *et al.* Protein quantitation through targeted mass spectrometry: the way out of biomarker purgatory? *Clin Chem.* 2008; 54(11): 1749-52.

Cazares, L. H., *et al.* Imaging mass spectrometry of a specific fragment of mitogen-activated protein kinase/extracellular signal-regulated kinase kinase kinase 2 discriminates cancer from uninvolved prostate tissue. *Clin Cancer Res.* 2009; 15(17): 5541-51.

Chen, G., *et al.* Qualitative and quantitative analysis of peptide microarray binding experiments using SVM-PEPARRAY. *Methods Mol Biol.* 2009; 570: 403-11.

Cho, W., *et al.* Use of glycan targeting antibodies to identify cancer-associated glycoproteins in plasma of breast cancer patients. *Anal Chem.* 2008; 80(14): 5286-92.

Choi, H., *et al.* False discovery rates and related statistical concepts in mass spectrometry-based proteomics. *J Proteome Res.* 2008; 7(1): 47-50.

Choi, H., *et al.* Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. *J Proteome Res.* 2008; 7(1): 254-65.

Choi, H., *et al.* Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling. *J Proteome Res.* 2008; 7(1): 286-92.

Choi, H., *et al.* Significance analysis of spectral count data in label-free shotgun proteomics. *Mol Cell Proteomics.* 2008; 7(12): 2373-85.

Clowers, B. H., *et al.* Enhanced ion utilization efficiency using an electrodynamic ion funnel trap as an injection mechanism for ion mobility spectrometry. *Anal Chem.* 2008; 80(3): 612-23.

Clowers, B. H., *et al.* Pseudorandom sequence modifications for ion mobility orthogonal time-of-flight mass spectrometry. *Anal Chem.* 2008; 80(7): 2464-73.

Dalkic, E., *et al.* Integrative Analysis of Cancer Pathway Progression. *Proteomics Clin. Appl.* 2009; 3: 473-85.

Ding, Y., *et al.* Adaptive discriminant function analysis and reranking of MS/MS database search results for improved peptide identification in shotgun proteomics. *J Proteome Res.* 2008; 7(11): 4878-89.

Drake, P. M., *et al.* Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation. *Clinical Chemistry.* 2009: in press.

Du, J., *et al.* Bead-based profiling of tyrosine kinase phosphorylation identifies SRC as a potential target for glioblastoma therapy. *Nat Biotechnol.* 2009; 27(1): 77-83.

Edwards, N. J. Novel peptide identification from tandem mass spectra using ESTs and sequence database compression. *Mol Syst Biol.* 2007; 3: 102.

Edwards, N., *et al.* An Unsupervised, Model-Free, Machine-Learning Combiner for Peptide Identifications from Tandem Mass Spectra. *Clinical Proteomics.* 2009; 5(1): 23-36.

Fedulova, I., *et al.* PepTiger: Search Engine for Error-Tolerant Protein Identification from de Novo Sequences. *The Open Spectroscopy Journal.* 2007; 1: 1-8.

Feng, J., *et al.* Peptide conformations for a microarray surface-tethered epitope of the tumor suppressor p53. *J Phys Chem B.* 2007; 111(49): 13797-806.

Feng, J., *et al.* Salt effects on surface-tethered peptides in solution. *J Phys Chem B.* 2009; 113(28): 9472-8.

Fenyo, D., *et al.* Determining the overall merit of protein identification data sets: rho-diagrams and rho-scores. *J Proteome Res.* 2007; 6(5): 1997-2004.

Fusaro, V. A., *et al.* Prediction of high-responding peptides for targeted protein assays by mass spectrometry. *Nat Biotechnol.* 2009; 27(2): 190-8.

Gatlin-Bunai, C. L., *et al.* Optimization of MALDI-TOF MS detection for enhanced sensitivity of affinity-captured proteins spanning a 100 kDa mass range. *J Proteome Res.* 2007; 6(11): 4517-24.

Gehrke, A., *et al.* Improved machine learning method for analysis of gas phase chemistry of peptides. *BMC Bioinformatics.* 2008; 9: 515.

Gloriam, D. E., *et al.* Report: A community standard format for the representation of protein affinity reagents. *Mol Cell Proteomics.* 2009.

Gong, W., *et al.* PepCyber:P~PEP: a database of human protein protein interactions mediated by phosphoprotein-binding domains. *Nucleic Acids Res.* 2008; 36(Database issue): D679-83.

Goudreault, M., *et al.* 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.* 2009; 8(1): 157-71.

Gredell, J. A., *et al.* Impact of target mRNA structure on siRNA silencing efficiency: A large-scale study. *Biotechnol Bioeng.* 2008; 100(4): 744-55.

Gredell, J. A., *et al.* (2009). Engineering Cell Function by RNA Interference. Cell Line Development, Springer Netherlands. **6**: 175-194.

Ho, L., *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A.* 2009; 106(13): 5181-6.

Holinger, E. P., *et al.* Budding yeast centrosome duplication requires stabilization of Spc29 via Mps1-mediated phosphorylation. *J Biol Chem.* 2009; 284(19): 12949-55.

Hortin, G. L., *et al.* High-abundance polypeptides of the human plasma proteome comprising the top 4 logs of polypeptide abundance. *Clin Chem.* 2008; 54(10): 1608-16.

Jaffe, J. D., *et al.* PEPPeR, a platform for experimental proteomic pattern recognition. *Mol Cell Proteomics.* 2006; 5(10): 1927-41.

Jaffe, J. D., *et al.* Accurate inclusion mass screening: a bridge from unbiased discovery to targeted assay development for biomarker verification. *Mol Cell Proteomics.* 2008; 7(10): 1952-62.

Johansen, E., *et al.* A lectin HPLC method to enrich selectively-glycosylated peptides from complex biological samples. *J Vis Exp.* 2009; (32).

Jung, K., *et al.* Glycoproteomics of plasma based on narrow selectivity lectin affinity chromatography. *J Proteome Res.* 2009; 8(2): 643-50.

Karbassi, I. D., *et al.* Proteomic expression profiling and identification of serum proteins using immobilized trypsin beads with MALDI-TOF/TOF. *J Proteome Res.* 2009; 8(9): 4182-92.

Keshishian, H., *et al.* Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics.* 2007; 6(12): 2212-29.

Keshishian, H., *et al.* Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol Cell Proteomics.* 2009; 8(10): 2339-49.

Kessner, D., *et al.* ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics.* 2008; 24(21): 2534-6.

Kim, B., *et al.* The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc Natl Acad Sci U S A.* 2007; 104(41): 16068-73.

Kini, H. K., *et al.* In vitro binding of single-stranded RNA by human Dicer. *FEBS Lett.* 2007; 581(29): 5611-6.

Kini, H. K., *et al.* Effect of siRNA terminal mismatches on TRBP and Dicer binding and silencing efficacy. *Febs J.* 2009.

Kuhn, E., *et al.* Developing multiplexed assays for troponin I and interleukin-33 in plasma by peptide immunoaffinity enrichment and targeted mass spectrometry. *Clin Chem.* 2009; 55(6): 1108-17.

Kumar, K. G., *et al.* Site-specific ubiquitination exposes a linear motif to promote interferon-alpha receptor endocytosis. *J Cell Biol.* 2007; 179(5): 935-50.

Kuzyk, M. A., *et al.* Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol Cell Proteomics.* 2009; 8(8): 1860-77.

Lee, B. T., *et al.* Candidate List of yoUr Biomarker (CLUB): A Web-based Platform to Aid Cancer Biomarker Research. *Biomark Insights.* 2008; 3: 65-71.

Li, Z., *et al.* Using dynamic gene module map analysis to identify targets that modulate free fatty acid induced cytotoxicity. *Biotechnol Prog.* 2008; 24(1): 29-37.

Li, Z., *et al.* Systems biology for identifying liver toxicity pathways. *BMC Proc.* 2009; 3 Suppl 2: S2.

Li, T., *et al.* Web-based design of peptide microarrays using microPepArray Pro. *Methods Mol Biol.* 2009; 570: 391-401.

Li, Y. F., *et al.* A bayesian approach to protein inference problem in shotgun proteomics. *J Comput Biol.* 2009; 16(8): 1183-93.

Li, J., *et al.* Network-assisted protein identification and data interpretation in shotgun proteomics. *Mol Syst Biol.* 2009; 5: 303.

Li, J., *et al.* Discovering Novel Breast Cancer Drug Candidates from Biomedical Literature. *International Journal of Data Mining and Bioinformatics.* 2009: in press.

Liebler, D. C. Summary of United States Human Proteome Organisation (HUPO) symposium entitled "Standardized clinical proteomics platforms". *Mol Cell Proteomics.* 2009; 8(5): 1165-6.

Liu, Z., *et al.* Human BRE1 is an E3 ubiquitin ligase for Ebp1 tumor suppressor. *Mol Biol Cell.* 2009; 20(3): 757-68.

Lu, Y., *et al.* USP19 deubiquitinating enzyme supports cell proliferation by stabilizing KPC1, a ubiquitin ligase for p27Kip1. *Mol Cell Biol.* 2009; 29(2): 547-58.

Ma, Z. Q., *et al.* IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *J Proteome Res.* 2009; 8(8): 3872-81.

Marelli, M., *et al*. Identifying bona fide components of an organelle by isotope-coded labeling of subcellular fractions : an example in peroxisomes. *Methods Mol Biol*. 2008; 432: 357-71.

Mattison, C. P., *et al*. Mps1 activation loop autophosphorylation enhances kinase activity. *J Biol Chem*. 2007; 282(42): 30553-61.

McConnell, R. E., *et al*. The enterocyte microvillus is a vesicle-generating organelle. *J Cell Biol*. 2009; 185(7): 1285-98.

McLerran, D., *et al*. Analytical validation of serum proteomic profiling for diagnosis of prostate cancer: sources of sample bias. *Clin Chem*. 2008; 54(1): 44-52.

McLerran, D., *et al*. SELDI-TOF MS whole serum proteomic profiling with IMAC surface does not reliably detect prostate cancer. *Clin Chem*. 2008; 54(1): 53-60.

McLerran, D. F., *et al*. Signal detection in high-resolution mass spectrometry data. *J Proteome Res*. 2008; 7(1): 276-85.

Mehrotra, S., *et al*. Multilayer mediated forward and patterned siRNA transfection using linear-PEI at extended N/P ratios. *Acta Biomater*. 2009; 5(5): 1474-88.

Mehrotra, S., *et al*. Time controlled protein release from layer-by-layer assembled multilayers functionalized agarose hydrogels. *Advanced Functional Materials*. 2009: in press.

Mueller, D., *et al*. A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. *Blood*. 2007; 110(13): 4445-54.

Mueller, L. N., *et al*. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J Proteome Res*. 2008; 7(1): 51-61.

Nesvizhskii, A. I. Protein identification by tandem mass spectrometry and sequence database searching. *Methods Mol Biol*. 2007; 367: 87-119.

Nesvizhskii, A. I., *et al*. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods*. 2007; 4(10): 787-97.

Oh, C., *et al*. Neural network prediction of peptide separation in strong anion exchange chromatography. *Bioinformatics*. 2007; 23(1): 114-8.

Old, W. M., *et al*. Functional proteomics identifies targets of phosphorylation by B-Raf signaling in melanoma. *Mol Cell*. 2009; 34(1): 115-31.

Ong, S. E., *et al*. Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc Natl Acad Sci U S A*. 2009; 106(12): 4617-22.

Ott, L. W., *et al*. Tumor Necrosis Factor-alpha- and interleukin-1-induced cellular responses: coupling proteomic and genomic information. *J Proteome Res*. 2007; 6(6): 2176-85.

Padliya, N. D., *et al*. Tandem mass spectrometry for the detection of plant pathogenic fungi and the effects of database composition on protein inferences. *Proteomics*. 2007; 7(21): 3932-42.

Page, J. S., *et al*. Biases in Ion Transmission Through an Electrospray Ionization-Mass Spectrometry Capillary Inlet. *J Am Soc Mass Spectrom*. 2009.

Paulovich, A. G., *et al*. The Interface between Biomarker Discovery and Clinical Validation: The Tar Pit of the Protein Biomarker Pipeline. *Proteomics Clin. Appl*. 2008; 2: 1386-1402.

Paulovich, A. G., *et al*. A CPTAC inter-laboratory study characterizing a yeast performance standard for benchmarking LC-MS Platform performance. *Mol Cell Proteomics*. 2009.

Peng, I. X., *et al*. Electrospray-assisted laser desorption/ionization and tandem mass spectrometry of peptides and proteins. *Rapid Commun Mass Spectrom*. 2007; 21(16): 2541-6.

Peng, I. X., *et al*. Reactive-electrospray-assisted laser desorption/ionization for characterization of peptides and proteins. *Anal Chem*. 2008; 80(18): 6995-7003.

Peterson, A., *et al*. Analysis of RP-HPLC loading conditions for maximizing peptide identifications in shotgun proteomics. *J Proteome Res*. 2009; 8(8): 4161-8.

Pope, M. E., *et al*. Anti-peptide antibody screening: selection of high affinity monoclonal reagents by a refined surface plasmon resonance technique. *J Immunol Methods*. 2009; 341(1-2): 86-96.

Qiu, R., *et al*. Comparative glycoproteomics of N-linked complex-type glycoforms containing sialic acid in human serum. *Anal Chem*. 2005; 77(22): 7225-31.

Qiu, R., *et al*. Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal Chem*. 2005; 77(9): 2802-9.

Qiu, R., *et al*. 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. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2007; 845(1): 143-50.

Qiu, J., *et al*. Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. *J Clin Oncol*. 2008; 26(31): 5060-6.

Ramesh Babu, J., *et al*. Genetic inactivation of p62 leads to accumulation of hyperphosphorylated tau and neurodegeneration. *J Neurochem*. 2008; 106(1): 107-20.

Riley, C. P., *et al*. The proteome discovery pipeline - a data analysis pipeline for mass spectrometry-based differential proteomics. *BMC Bioinformatics*. 2009: in press.

Ritchie, J., *et al*. Proteomics - Is There a Role for Clinical Labs Now? *American Association for Clinical Chemistry - Clinical Laboratory News*. 2009; 35(2): 12-14.

Rodriguez, H. International summit on proteomics data release and sharing policy. *J Proteome Res*. 2008; 7(11): 4609.

Rodriguez, H., *et al*. Recommendations from the 2008 International Summit on Proteomics Data Release and Sharing Policy: the Amsterdam principles. *J Proteome Res*. 2009; 8(7): 3689-92.

Rodriguez, H. Complimentary Fields of Genomics and Proteomics are Cornerstone of Personalized Medicine. *Drug Discovery News*. 2009; (March): 13.

Rodriguez, H. Proteomics: A Long and Winding Road to Medical Diagnostics. *Expert Opinion on Medical Diagnostics*. 2009; 3(3): 219-225.

Rodriguez, H. Restructuring Proteomics to Enable Personalized Cancer Care. *Drug Discovery World*. 2009; (August): 17-22.

Rudnick, P. A., *et al*. Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomic analyses and evaluation by the CPTAC network. *Mol Cell Proteomics*. 2009.

Rudomin, E. L., *et al*. Directed sample interrogation utilizing an accurate mass exclusion-based data-dependent acquisition strategy (AMEx). *J Proteome Res*. 2009; 8(6): 3154-60.

Saha, S., *et al*. HIP2: an online database of human plasma proteins from healthy individuals. *BMC Med Genomics*. 2008; 1: 12.

Saha, S., *et al*. Dissecting the human plasma proteome and inflammatory response biomarkers. *Proteomics*. 2009; 9(2): 470-84.

Searle, B. C., *et al*. Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. *J Proteome Res*. 2008; 7(1): 245-53.

Seyfried, N. T., *et al*. Systematic approach for validating the ubiquitinated proteome. *Anal Chem*. 2008; 80(11): 4161-9.

Shen, C., *et al*. A hierarchical statistical model to assess the confidence of peptides and proteins inferred from tandem mass spectrometry. *Bioinformatics*. 2008; 24(2): 202-8.

Sherwood, C. A., *et al*. Rapid optimization of MRM-MS instrument parameters by subtle alteration of precursor and product m/z targets. *J Proteome Res*. 2009; 8(7): 3746-51.

Sherwood, C. A., *et al*. MaRiMba: a software application for spectral library-based MRM transition list assembly. *J Proteome Res*. 2009; 8(10): 4396-405.

Sherwood, C. A., *et al*. Correlation between y-type ions observed in ion trap and triple quadrupole mass spectrometers. *J Proteome Res*. 2009; 8(9): 4243-51.

Slebos, R. J., *et al*. Evaluation of strong cation exchange versus isoelectric focusing of peptides for multidimensional liquid chromatography-tandem mass spectrometry. *J Proteome Res*. 2008; 7(12): 5286-94.

Sprung, R. W., Jr., *et al*. 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*. 2009; 8(8): 1988-98.

Srivastava, S., *et al*. Identification of genes that regulate multiple cellular processes/responses in the context of lipotoxicity to hepatoma cells. *BMC Genomics*. 2007; 8: 364.

Srivastava, S., *et al*. A novel method incorporating gene ontology information for unsupervised clustering and feature selection. *PLoS One*. 2008; 3(12): e3860.

Sun, S., *et al*. Improved validation of peptide MS/MS assignments using spectral intensity prediction. *Mol Cell Proteomics*. 2007; 6(1): 1-17.

Swatkoski, S., *et al*. Integration of residue-specific acid cleavage into proteomic workflows. *J Proteome Res*. 2007; 6(11): 4525-7.

Swatkoski, S., *et al*. Evaluation of microwave-accelerated residue-specific acid cleavage for proteomic applications. *J Proteome Res*. 2008; 7(2): 579-86.

Tabb, D. L., *et al*. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res*. 2007; 6(2): 654-61.

Tabb, D. L. What's driving false discovery rates? *J Proteome Res*. 2008; 7(1): 45-6.

Tabb, D. L., *et al*. DirecTag: accurate sequence tags from peptide MS/MS through statistical scoring. *J Proteome Res*. 2008; 7(9): 3838-46.

Taylor, B. S., *et al*. Humoral response profiling reveals pathways to prostate cancer progression. *Mol Cell Proteomics*. 2008; 7(3): 600-11.

Taylor, C. F., *et al.* Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol*. 2008; 26(8): 889-96.

Tharakan, R., *et al.* Data Maximization by Multipass Analysis of Protein Mass Spectra. *Proteomics*. 2010: in press.

Tracy, M. B., *et al.* Precision enhancement of MALDI-TOF MS using high resolution peak detection and label-free alignment. *Proteomics*. 2008; 8(8): 1530-8.

Ulintz, P. J., *et al.* Investigating MS2/MS3 matching statistics: a model for coupling consecutive stage mass spectrometry data for increased peptide identification confidence. *Mol Cell Proteomics*. 2008; 7(1): 71-87.

Ulintz, P. J., *et al.* Comparison of MS(2)-only, MSA, and MS(2)/MS(3) methodologies for phosphopeptide identification. *J Proteome Res*. 2009; 8(2): 887-99.

Varghese, B., *et al.* Polyubiquitination of prolactin receptor stimulates its internalization, postinternalization sorting, and degradation via the lysosomal pathway. *Mol Cell Biol*. 2008; 28(17): 5275-87.

Vellaichamy, A., *et al.* Proteomic interrogation of androgen action in prostate cancer cells reveals roles of aminoacyl tRNA synthetases. *PLoS One*. 2009; 4(9): e7075.

Victor, K. G., *et al.* MAZIE: A Mass and Charge Inference Engine to Enhance Database Searching of Tandem Mass Spectra. *J Am Soc Mass Spectrom*. 2009.

Villanueva, J., *et al.* Serum peptidome patterns that distinguish metastatic thyroid carcinoma from cancer-free controls are unbiased by gender and age. *Mol Cell Proteomics*. 2006; 5(10): 1840-52.

Villanueva, J., *et al.* Automated serum peptide profiling. *Nat Protoc.* 2006; 1(2): 880-91.

Villanueva, J., *et al.* Data analysis of assorted serum peptidome profiles. *Nat Protoc*. 2007; 2(3): 588-602.

Villanueva, J., *et al.* A sequence-specific exopeptidase activity test (SSEAT) for "functional" biomarker discovery. *Mol Cell Proteomics*. 2008; 7(3): 509-18.

Villanueva, J., *et al.* Monitoring peptidase activities in complex proteomes by MALDI-TOF mass spectrometry. *Nat Protoc*. 2009; 4(8): 1167-83.

Vitzthum, F., *et al.* 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*. 2007; 1: 1016-1035.

Walton, S. P., *et al*. Proteomics: technology development and applications. *Expert Rev Proteomics*. 2009; 6(1): 23-5.

Wan, J., *et al.* Meta-prediction of phosphorylation sites with weighted voting and restricted grid search parameter selection. *Nucleic Acids Res*. 2008; 36(4): e22.

Wang, J., *et al.* Integration of 18O labeling and solution isoelectric focusing in a shotgun analysis of mitochondrial proteins. *J Proteome Res*. 2007; 6(12): 4601-7.

Wang, X., *et al.* Gene module level analysis: identification to networks and dynamics. *Curr Opin Biotechnol*. 2008; 19(5): 482-91.

Wang, W., *et al*. Structural and Mechanistic Insights into Mps1 Kinase Activation. *J Cell Mol Med.* 2008.

Wang, P., *et al.* The evolving role of mass spectrometry in cancer biomarker discovery. *Cancer Biol Ther.* 2009; 8(12): 1083-94.

Whiteaker, J. R., *et al*. Antibody-based enrichment of peptides on magnetic beads for mass-spectrometry-based quantification of serum biomarkers. *Anal Biochem*. 2007; 362(1): 44-54.

Whiteaker, J. R., *et al*. Integrated pipeline for mass spectrometry-based discovery and confirmation of biomarkers demonstrated in a mouse model of breast cancer. *J Proteome Res*. 2007; 6(10): 3962-75.

Whiteaker, J. R., *et al*. An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. *Mol Cell Proteomics*. 2009.

Williams, B. A., *et al.* Creating Protein Affinity Reagents by Combining Peptide Ligands on Synthetic DNA Scaffolds. *J Am Chem Soc*. 2009.

Winborn, B. J., *et al.* The deubiquitinating enzyme ataxin-3, a polyglutamine disease protein, edits Lys63 linkages in mixed linkage ubiquitin chains. *J Biol Chem*. 2008; 283(39): 26436-43.

Witze, E. S., *et al.* Mapping protein post-translational modifications with mass spectrometry. *Nat Methods*. 2007; 4(10): 798-806.

Wooten, M. W., *et al.* Essential role of sequestosome 1/p62 in regulating accumulation of Lys63-ubiquitinated proteins. *J Biol Chem*. 2008; 283(11): 6783-9.

Wu, X., *et al.* HMMatch: peptide identification by spectral matching of tandem mass spectra using hidden Markov models. *J Comput Biol*. 2007; 14(8): 1025-43.

Wu, M., *et al.* A dynamic analysis of insulin signaling and its feedback mechanisms: A discrete modeling approach. *PLoS ONE*. 2009: in press.

Wynne, C., *et al.* Top-Down Identification of Protein Biomarkers in Bacteria with Unsequenced Genomes. *Anal Chem*. 2009.

Xia, Y., *et al.* A comprehensive sequence and disease correlation analyses for the C-terminal region of CagA protein of Helicobacter pylori. *PLoS One*. 2009; 4(11): e7736.

Xie, S., *et al.* Application and analysis of structure-switching aptamers for small molecule quantification. *Anal Chim Acta.* 2009; 638(2): 213-9.

Xiong, L., *et al.* Comparative proteomics of glycoproteins based on lectin selection and isotope coding. *J Proteome Res*. 2003; 2(6): 618-25.

Xu, P., *et al.* Characterization of polyubiquitin chain structure by middle-down mass spectrometry. *Anal Chem.* 2008; 80(9): 3438-44.

Xu, Q., *et al.* Regulation of kinetochore recruitment of two essential mitotic spindle checkpoint proteins by Mps1 phosphorylation. *Mol Biol Cell.* 2009; 20(1): 10-20.

Xu, P., *et al.* Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell.* 2009; 137(1): 133-45.

Yang, X., *et al.* Repression of PKR mediates palmitate-induced apoptosis in HepG2 cells through regulation of Bcl-2. *Cell Res.* 2009; 19(4): 469-86.

Yang, X., *et al*. Reconstruct modular phenotype-specific gene networks by knowledge-driven matrix factorization. *Bioinformatics*. 2009; 25(17): 2236-43.

Yang, X., *et al.* (2009). Construction of the gene network by synergy analysis of the genes related to palmitate-induced cytotoxicity. Methods in Bioengineering: Systems Analysis of Biological Networks. A. Jayaraman and J. Hahn, Artech House Publishers: 75 - 93.

Yen, C. Y., *et al.* A simulated MS/MS library for spectrum-to-spectrum searching in large scale identification of proteins. *Mol Cell Proteomics.* 2009; 8(4): 857-69.

Yoo, C., *et al.* Comprehensive analysis of proteins of pH fractionated samples using monolithic LC/MS/MS, intact MW measurement and MALDI-QIT-TOF MS. *J Mass Spectrom*. 2007; 42(3): 312-34.

Yu, X., *et al.* Detecting Genomic Aberrations Using Products in a Multiscale Analysis. *Biometrics*. 2009.

Zhang, B., *et al.* Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J Proteome Res*. 2007; 6(9): 3549-57.

Zhang, X., *et al.* Current Status of Computational Approaches for Protein Identification Using Tandem Mass Spectra. *Current Proteomics*. 2007; 4(3): 121-130.

Zhang, M., *et al.* Interactive analysis of systems biology molecular expression data. *BMC Syst Biol.* 2008; 2: 23.

Zhu, Q., *et al.* microParaflo biochip for nucleic acid and protein analysis. *Methods Mol Biol*. 2007; 382: 287-312.


NATIONAL CANCER INSTITUTE

**NIH Publication No. 10 7508**
**Printed December 2009**