

# GenoLogics Case Study: The Bindley Bioscience Center, Discovery Park at Purdue University Validating the Cancer Biomarker Supply Chain



## Case Study

In September 2006, the Analytical Proteomics Team from Purdue and Indiana universities was selected by the National Cancer Institute (NCI) as one of five national centers to assess cancer biomarker proteomics technologies. The APT, along with Broad Institute of MIT and Harvard, University of California at San Francisco Lawrence Berkeley National Laboratory, Memorial Sloan-Kettering Cancer Center, and Vanderbilt University School of Medicine formed the Clinical Proteomics Technology Assessment for Cancer (CPTAC) network. The CPTAC program is a five-year, \$104 million NCI initiative to create a foundation of technologies, data, reagents and standards, analysis systems, and infrastructure for the discovery and validation of clinically relevant cancer biomarkers. The consortium is focused on systematically advancing understanding of protein biology in cancer, and accelerating discovery research and clinical applications.

Purdue University's Bindley Bioscience Center at Discovery Park is the Analytical Proteomics Team's base for developing protocols and standards for mass spectrometry-based cancer proteomics relating to breast and prostate cancer, research that involves collaboration among experts in proteomics, informatics, and cancer biology and treatment.

The team at the Bindley Bioscience Center is currently evaluating existing proteomic analysis platforms to reliably identify, quantify, and compare peptides and proteins in complex biological mixtures, using a variety of mass spectrometry and affinity-based technologies, in a multidisciplinary team approach. In addition to both electrospray ionization and matrix assisted laser desorption ionization mass spectrometry platform, the team is also using 'bioCD' technology, a new microarray technology invented at Purdue that employs interferometric analysis for the detection and quantification of specific candidate cancer protein biomarkers.

The Bindley Bioscience Center's CPTAC work involves significant informatics and data handling challenges. The Analytical Proteomics Team must meet rigorous data sharing requirements, achieve caBIG compliance, establish data communication, and develop models for data handling and sharing between sites. In order to address these challenges, Purdue University requires effective tools for data capture, analysis, and knowledge creation. The Bindley Bioscience Center began using GenoLogics' Proteus lab and data management solution in 2006, and Proteus' robust LIMS capabilities have facilitated linking, storing, and analyzing data with computational and statistical tools.

The Bindley Bioscience Center has developed a variety of tools that distinguish it among proteomics researchers. Purdue's CPTAC research is focused on locating differences between disease (cancer) and control samples, and then digging deeper into these results to find low abundant proteins, eventually resulting in knowledge that aids biomarker discovery.

The Bindley Bioscience Center has developed a unique experimental design (figure 1), spearheaded by Xiang Zhang, lead scientist for bioinformatics, that allows deeper analysis than that permitted by conventional workflows to find low abundant proteins.

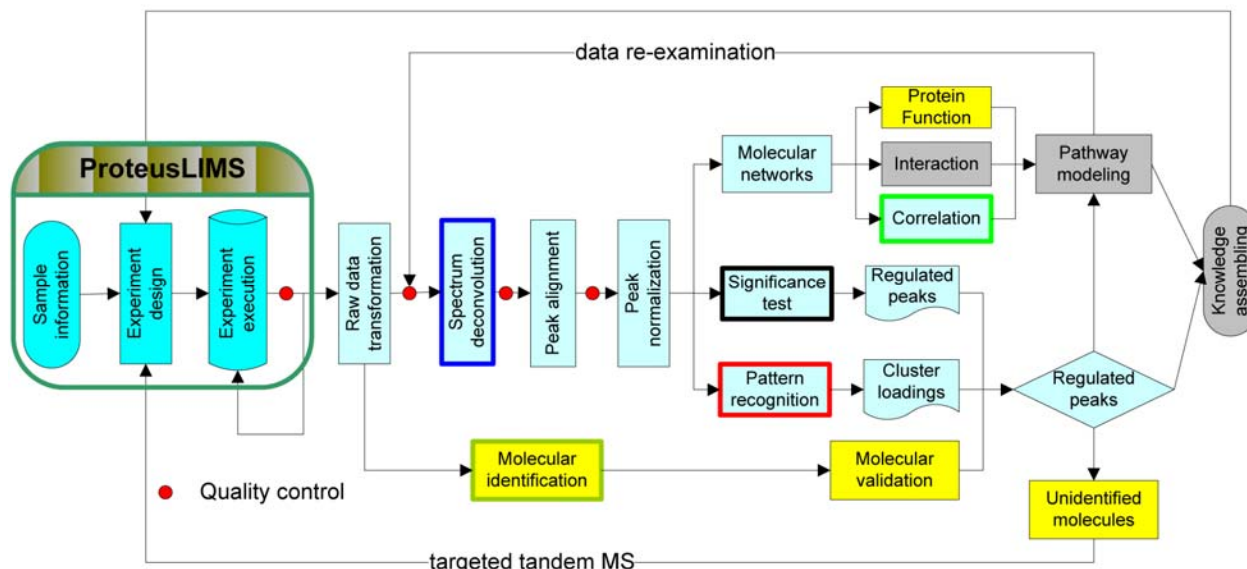


Figure 1. Bindley Bioscience Center experimental design

This unique experiment design includes three processes. First, a randomized selection of control and disease samples is subjected to LC-MS. For quality control, a pooled sample is analyzed in full MS mode following every five biological samples. A fraction of samples from control and disease samples are randomly selected for LC-MS/MS analysis by selecting the top three abundant ions in each full MS scan. Data from the Bindley's MALDI TOF TOF and other mass spectrometers are automatically gathered by Proteus, which then seamlessly pipes the data back through Purdue's pipeline (figure 1). The results of each analysis step are then zipped into a single file and sent back into Proteus.

The second component of the experiment design involves conducting MS/MS analyses to identify a group of significantly regulated peaks. Low abundance proteins are identified following alignment with previous LC-MS/MS data to find regulated peaks for which MS/MS information has not yet been collected. Finally, the researchers conduct targeted MS/MS analysis on the samples with significant peaks from the first phase.

In isolating these techniques, the Bindley Bioscience Center has created a distinct proteomics workflow. Most proteomics research involves identification and quantification in a single experiment (MS/MS), but Dr. Zhang's lab performs MS/MS on a selection of samples, and then further processes the data to identify the low abundance proteins, resulting in more efficient knowledge assembly.

GenoLogics supports Purdue’s novel research methods with an open, science-purposed bioinformatics platform. Several configurable processes were created within Proteus to seamlessly integrate the data generated by Dr. Zhang’s unique software directly into an end-to-end sample workflow. Dr. Zhang’s custom-built software includes XMass<sup>1</sup> and XAlign<sup>2</sup>, which deconvolutes spectra with isotopic peak profiles and aligns peptide peaks generated in each biological sample, and SysNet<sup>3</sup> (figure 2), which assists in correlation analysis by providing interactive visual data mining. Dr. Zhang’s team is also developing a tool for analyses of multiple protein search results with machine learning<sup>4</sup>.

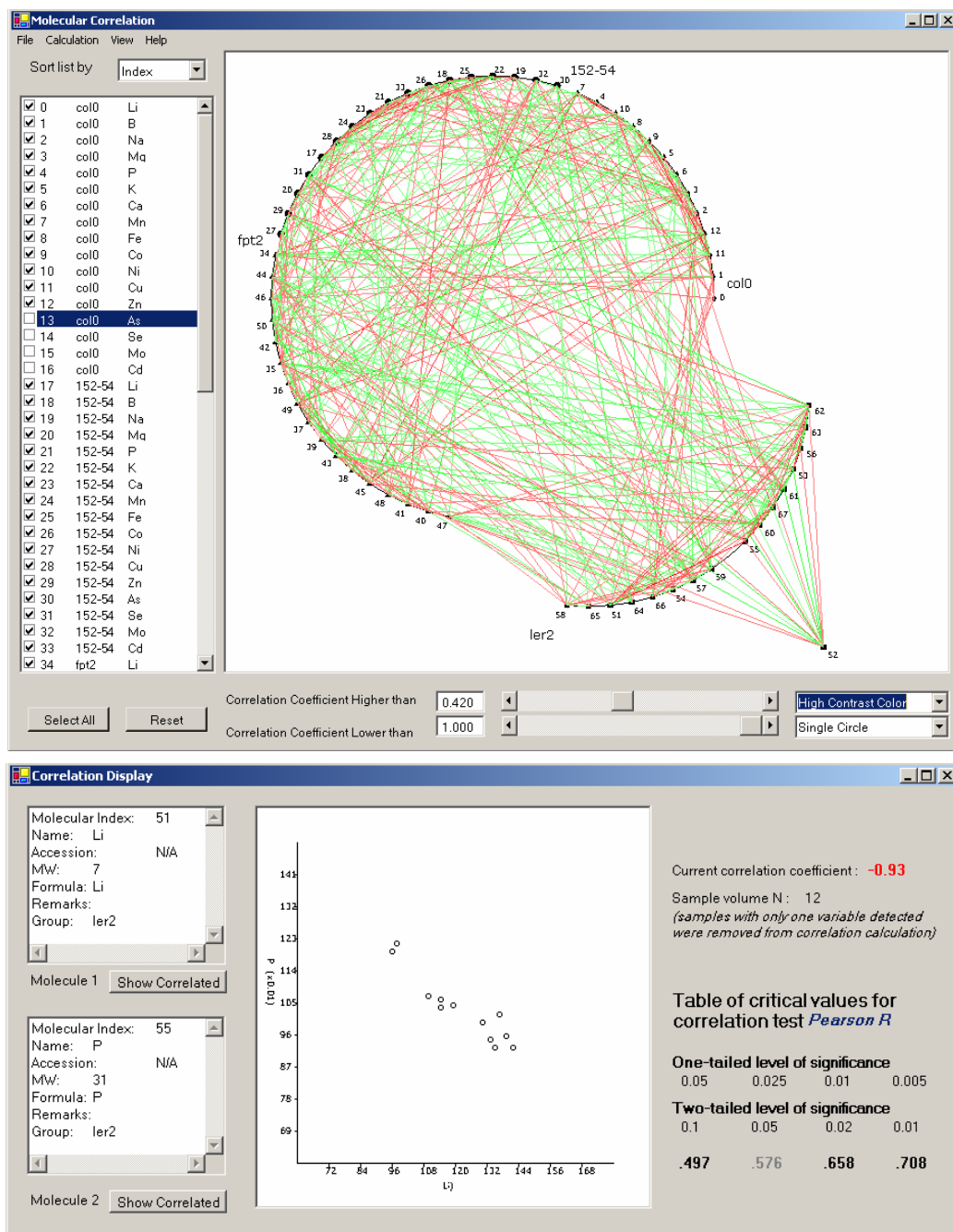


Figure 2. SysNet, developed by Purdue University to provide interactive visual data mining

In addition to integrating to leading commercial software and instruments, Proteus also easily configures to Purdue's specialized, unique tools. This flexibility has supported the Bindley Bioscience Center's involvement in CPTAC, and has contributed to Purdue University's overall goal of developing and promoting predictive, personalized, preventive healthcare.

---

<sup>1</sup> XMass

Zhang, X.; Hines, W.; Adamec, J.; Asara, J.; Naylor, S. and Regnier, F. E. An automated method for the analysis of stable isotope labeling data for proteomics. *J. Am. Soc. Mass Spectrom.* 2005, 16, 1181-1191.

<sup>2</sup> XAlign

Zhang, X.; Asara, J. M.; Adamec, J.; Ouzzani, M. and Elmagarmid, A. K. Data preprocessing in liquid chromatography mass spectrometry based proteomics. *Bioinformatics*, 2005, 21, 4054-4059.

<sup>3</sup> SysNet

Zhang, M.; Ouyang, Q.; Stephenson, A.; Kane, M. D.; Salt, D. E.; Prabhakar, S.; Buck, C. and Zhang, X. Interactive analysis of 'omics molecular expression data. *BMC Bioinformatics*, submitted.

<sup>4</sup> ANN Prediction

Oh, C.; Zak, S. H.; Mirzaei, H.; Regnier, F. E. and Zhang, X. Neural network prediction of peptide behavior in strong anion exchange chromatography. *Bioinformatics*, 2007, 23, 114-118.