

Mass Spectrometry in Cancer Diagnosis – A Case Study of the INCOGEN VIBE Workflow Software

Background

Recent advances in genomics and proteomics have greatly increased our understanding of the molecular basis for functions of organisms. However, the characterization of single genes or proteins has provided only limited insight and benefits toward early diagnoses, improved subtyping and prognoses, and treatment of diseases such as cancer. To understand the intricate web of interactions that makes up the biological functioning of life, we must try to decipher how a gene or protein fits into this dynamic environment with thousands of other genes and proteins. The interpretation of these dynamic systems is vastly more complex than, for example, sequencing the human genome, which is linear and, for the purpose of sequencing, static. A comprehensive understanding of biological phenomena can only be achieved through the melding of information and insights from technologies that characterize genes and proteins at the level of sequence, transcription, regulation, structure, function, kinetics, and localization.

This integration of knowledge requires a departure from conventional approaches toward life science research and is only possible by combining state-of-the-art technologies and enabling knowledge exchange from traditionally divergent fields such as molecular biology, clinical research, computational science, physics, statistics, and hardware engineering (NIH Roadmap, Zerhouni 2003). This can only be achieved in an environment that enables meaningful and efficient integration of knowledge and technology from several different fields. At the heart of this environment must stand a sophisticated *bioinformatics framework* that allows researchers to meld their distinct expertise and most efficiently apply their contributions toward the common goal.

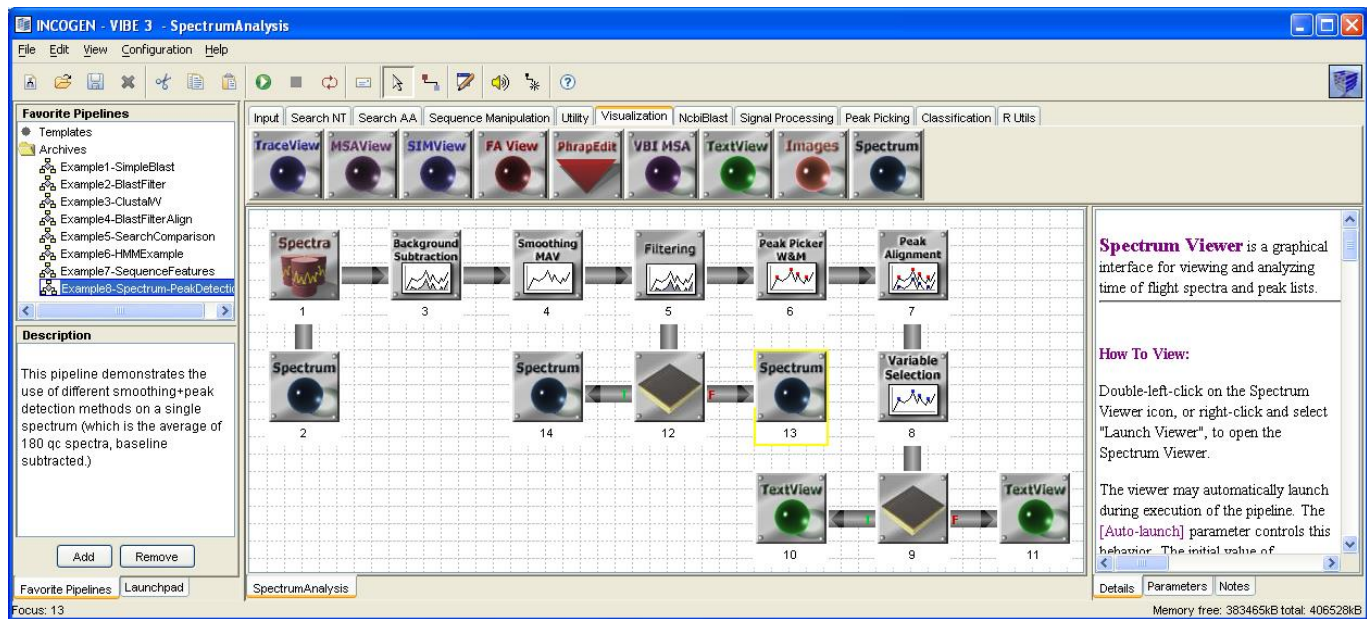
Specifically, the use of mass spectrometry for clinical applications has extraordinary potential for accurate, early, and minimally invasive diagnoses of complex diseases, such as cancer, which require sensitive diagnostic tools for prognosis and development of flexible treatment strategies. However, most mass spectrometry data analysis options available to researchers often require *ad hoc* combinations of various methods that are inefficient, repetitive and, in general, fall short of allowing researchers to fully exploit the data available to them. This greatly limits the researchers' ability to turn the information contained in the data into knowledge and associated benefits.

Bioinformatics Workflow Management Using the INCOGEN VIBE Software

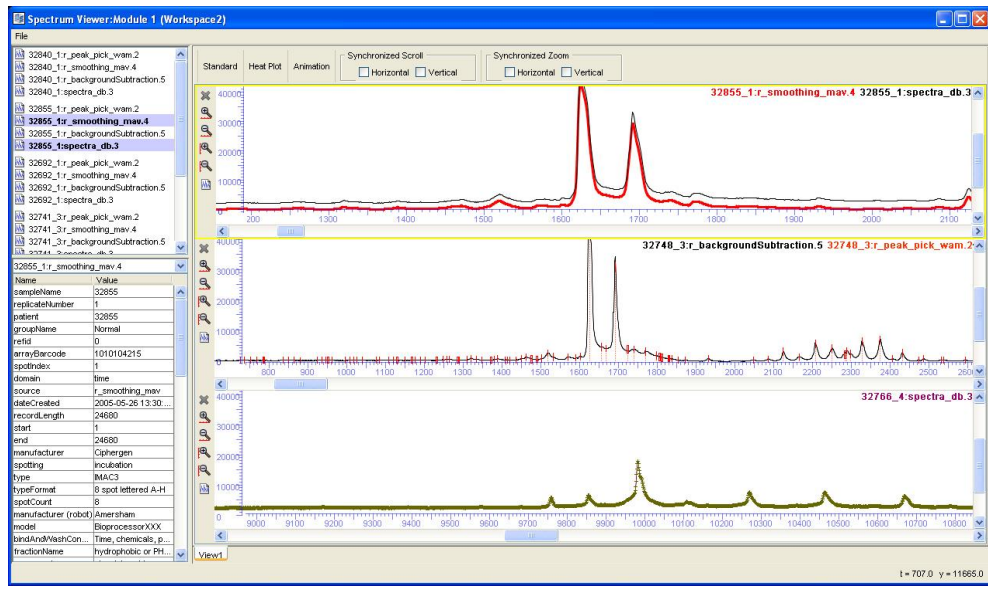
The INCOGEN VIBE (Visual Integrated Bioinformatics Environment) platform enables researchers to easily incorporate and access diverse, distributed tools and combine them into analysis workflows. VIBE consists of a scalable, extensible, multi-tiered workflow system designed as an integration platform for algorithms written in most programming languages and executable on most operating systems. The VIBE system can run on a single workstation, provided it is sufficiently powerful to support the included toolkits. It can also be distributed across multiple computing resources to fully exploit the IT support available to the users. The software has been recognized as the leading workflow management software for life sciences and was inducted into the ComputerWorld "Hall of Fame" in 2003. The ComputerWorld Global Archives, which reside in more than 125 of the leading museums, archives, libraries and academic institutions in 35 countries, recognize leading examples of the use of information technology to benefit society.

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. The ease with which modules can be removed from a pipeline, replaced with an alternate module, re-parameterized and used reduces the amount of time and effort required for a researcher to arrive at the optimal analysis protocol for a given problem.

The inherent interactivity and flexibility in managing and optimizing mass spectrometry analysis workflows proved of significant value to scientists during a recent prostate cancer study by INCOGEN researchers and their collaborators at The College of William and Mary and Eastern Virginia Medical School. Using the VIBE Mass Spectrometry toolkit, the researchers were able to: (1) quantitatively characterize instrument noise and determine optimal experimental protocols that enhance diagnostic sensitivity, (2) identify diagnostically significant features in the mass spectra and allow unequivocal association of those features with organ or cell-related disease outputs, and (3) select a classification scheme that optimized diagnostic classification rates (Malyarenko 2004, Semmes 2004, Malyarenko 2005, Chen 2005). This allowed researchers to maximize the knowledge gained from the data available to them to provide prognostic insight into the temporal or anatomical progression of the disease. Figures 1a and 1b present two screenshots of the VIBE Mass Spectrometry toolkit.



(a)



(b)

Figure 1: Example screenshots of VIBE-MS. (a) A mass spectrometry analysis pipeline. The icons represent analysis modules that can be dragged onto the main canvas to construct pipelines, or workflows. The modules represent tasks such as input of raw data from the experimental platform, preprocessing, selection of diagnostically significant variables, and classification. (b) An example screenshot of the VIBE mass spectrometry viewer.

References

Chen, H., Tracy, E.R., Cooke, W., Sasinowski, M., Trosset, M.: Bayesian Methods for Automated Peak Identification in Time-of-Flight Mass Spectroscopy. *In Preparation* (2005)

Malyarenko, D. I., Chen, H., Wilkerson, A.L., Tracy, E.R., Cooke, W.E., Manos, D.M., Sasinowski, M., Semmes, O.J.: Ga⁺ TOF-SIMS Lineshape Analysis for Resolution Enhancement of MALDI MS Spectra of a Peptide Mixture. *Appl. Surf. Sci.*, **231-232**:357 (2004)

Malyarenko D.I., Cooke W.E., Adam B.L., Malik G, Chen H, Teacy E.R., Trosset M.W., Sasinowski M, Semmes O.J., Manos D.M.: Enhancement of sensitivity and resolution of surface-enhanced laser desorption/ionization time-of-flight mass spectrometric records for serum peptides using time-series analysis techniques. *Clin. Chem.* **51(1)**: 1-2 (2005)

Semmes O.J., Feng Z, Adam B.L., Banez L.L., Bigbee W.L., Campos D, Cazares L.H., Chan D.W., Grizzle W.E., Izbicka E, Kagan J, Malik G, McLerran D, Moul J.W., Partin A, Prasanna P, Rosenzweig J, Sokoll L.J., Srivastava S, Thompson I, Welsh M.J., White N. Winget M, Yasui Y, Zhang Z, Zhu, L: Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin. Chem.* **51(1)**: 3-5 (2005)

Zerhouni, E.: The NIH Roadmap. *Science* **302**: 62-72 (2003)