

Poster G-21

The iProXpress Knowledge System for Proteomic Studies



Authors:

Hongzhan Huang (*Georgetown University Medical Center*)

Zhang-Zhi Hu (*Georgetown University Medical Center*)

Xin Yuan (*Georgetown University Medical Center*)

Peter McGarvey (*Georgetown University Medical Center*)

An Chi (*University of Virginia*)

Satya Saxena (*National Institute on Aging*)

Cathy H. Wu (*Georgetown University Medical Center*)

Short Abstract: iProXpress is an integrated system for proteomic and genomic analysis consisted of the PIR integrated protein information, analytical tools for sequence analysis and functional annotation, and a graphical user interface for categorization and visualization of expression data. This system has been successfully applied to several global profiling and functional analyses.

Long Abstract:

The iProXpress Knowledge System for Proteomic Studies

Hongzhan Huang¹ ; Zhang-Zhi Hu¹; Xin Yuan¹; Peter McGarvey¹; An Chi²; Satya Saxena³; Cathy H. Wu¹

¹Georgetown University Medical Center, Washington, DC, USA;

²University of Virginia, Charlottesville, VA 22904, USA;

³National Institute on Aging, Baltimore, MD 21224, USA

The accelerated growth of proteomics data presents both opportunities and challenges. Large-scale proteomic profiling of biological samples such as cells, organelles or biological fluids has led to the discovery of numerous key and novel proteins involved in many biological/disease processes including cancers, as well as to the identification of novel disease biomarkers and potential therapeutic targets. Advanced bioinformatics infrastructure and systems are needed for data integration and functional interpretation of large-scale proteomic data.

iProXpress (integrated Protein eXpression) is an integrated protein expression analysis system, which is designed to help analyze proteomic and genomic data such as protein/peptide and gene profiles from IP, 2D and MS proteomic and microarray gene expression experiments. The iProXpress knowledge system consists of three major components, the PIR data warehouse with integrated protein information, analytical tools for sequence analysis and functional annotation, and a graphical user interface for categorization and visualization of expression data. The system includes the following functionalities: 1) Gene/Peptide to Protein Mapping. Gene or protein lists are mapped to corresponding entries in UniProtKB of all known proteins based on gene/protein IDs, names or sequences. 2) Protein Information Matrix. Protein family, domain, and functional site

features for each protein are identified by BLAST, HMM, signal peptide, transmembrane helix predictions and other automated searches. For direct comparison of expressed genes/proteins, a comprehensive protein information matrix is generated, summarizing salient features retrieved from the underlying PIR data warehouse or inferred based on sequence similarity. 3) Protein Data Analysis for Pathway and Network Discovery. Users can conduct iterative categorization and sorting of proteins in the information matrix and correlate expression and interaction patterns to salient protein properties for pathway and network discovery. Proteins are clustered based on functions, pathways, and/or other attributes in the information matrix to identify hidden relationships not apparent in the data on expression patterns and interacting proteins, and to recognize candidate proteins of unknown identity that warrant further investigation. This global bioinformatics analysis provides a composite view of functional changes to help identify critical nodes and hidden relationships in the biological pathways and networks. The iterative categorization steps in the process are currently conducted manually; however, many of them can be automated and rules can be developed to flag significant clusters.

The system has been applied to several studies including the expression profile analysis of hormone-induced changes in endocrine tumor cells. It is also currently being adopted for analyses of pathogen/host genomic and proteomic data produced from the NIAID Biodefense Proteomic Program. Here we presented two case studies: 1) Global profiling of LC-MS proteome of human stem cell Ntera2. A total of about 9,000 distinct representative proteins were identified by iProXpress System. This large scale proteomic profiling of Ntera2 cell and its comparison with DNA microarray data would contribute to a better understanding of gene regulation at global level. 2) Analysis of organellar proteomes of various stages of melanosomes from human melanoma cell lines using the iProXpress knowledge systems, which included: Mapping to known mouse coat color genes led to identification of 17 essential human melanosome proteins; Identification of possible stage-specific melanosome proteins for validation; Comparison of melanosome proteome with those of several other organelles permitted a proposed list of proteins characteristic of melanosome. This study has greatly facilitated a better understanding of melanin synthesis and melanosome biogenesis pathways.