

*An Open House*  
**Center for Mathematical Biosciences  
& Center for Bio-Computing**

*Friday, October 26, 2007*

**ABSTRACTS**  
OF  
**CENTER FACULTY PRESENTATIONS**

## **Abstracts are arranged according to order of presentation**

### **Mass Spectrometry Based Proteomics and Computational Diagnostics: Profiling of Plasma Samples in Coronary Artery Disease**

Susanne Ragg, MD, PhD,<sup>b,c</sup> Ilka Ott, MD, PhD,<sup>d</sup> Vladimir Fokin, PhD,<sup>e</sup> Krzysztof Podgorski, PhD,<sup>e</sup>  
Olga Vitek, PhD,<sup>f</sup> Gunther Schadow, MD, PhD<sup>a,b,c</sup>

<sup>a</sup> *Indiana University School of Informatics, Indianapolis, IN*

<sup>b</sup> *Regenstrief Institute, Indianapolis, IN*

<sup>c</sup> *Indiana University School of Medicine, Indianapolis, IN*

<sup>d</sup> *Technical University Munich, Department of Medicine, Munich, Germany*

<sup>e</sup> *Department of Mathematical Sciences and Center for Mathematical Biosciences, IUPUI*

<sup>f</sup> *Department of Statistics, Purdue University, West Lafayette, IN*

In the United States, heart disease is the leading cause of death, accounting for 700,000 deaths per year. Several clinical studies of cardiovascular disease have conclusively shown that the combined use of multiple protein markers representing different pathophysiological processes improves risk stratification in patients with acute coronary syndromes. Thus there is an increased interest in integrating global proteomic profiling into clinical trials. However, transforming proteomic profiles into clinically relevant information requires complex computational methods for their analysis, integration with clinical data, and interpretation. Such computational methods are now routinely used by pharmaceutical companies such as Eli Lilly & Co, but as of yet they are not fully tuned to and adopted in academic medicine. As a result, publicly available analysis tools have not been tested on larger academic center-based studies and almost no information about global proteomic profiling in various diseases is available.

In our study blood was obtained from a total of 3500 patients seen at the Heart Center over a period of one year. Fifty patients, stratified for age and sex, were randomly selected from five subgroups. Plasma samples were processed for LC/MS/MS analysis using one Thermo-Finnigan linear ion-trap mass spectrometer. Overall 2581 proteins were identified and quantified. Several issues are currently being addressed to allow using high throughput proteomics data for clinical biomarker discovery.

Since protein quantities are derived from peptides, and since the same peptide could have been contributed from multiple different proteins, assigning the correct peptide quantities to the proteins needs to be accurate. To address this issue we re-analyze the data on the peptide level.

Prior information such as the known normal plasma concentration of proteins is used to answer the crucial question how far in the low-end of the abundance range the present technology can detect and quantify proteins.

The detected proteins are compared with a list of known markers in cardiovascular disease and a list of proteins detected in human plasma with high confidence by the HUPO Plasma Proteome Project. This will help avoid being distracted by spurious misidentifications and establish an accurate list of confidently identified proteins that are differentially expressed in the different stages of disease.

### **GeneTerrain: A Novel Visual Platform for Panel Biomarker Discoveries**

Qian You, Shiaofen Fang, PhD, and Jake Chen, PhD

*Department of Computer & Information Science and Center for Bio-Computing, IUPUI*

We invented a new visualization technique to explore and discover potential molecular biomarkers. Compared with heat map visualizations popular in bioinformatics, the GeneTerrain shows disease-related genes and proteins in their native protein interaction network neighborhood context, instead of an arbitrary matrix format. It integrates differential gene/protein expression values for each gene/protein with differential gene/protein expression values contributed from its close neighbors, and generates a final smooth surface to represent the end results. We show how this technique can be applied to Alzheimer's Disease data set, which we discovered some interesting early diagnosis and staging panel biomarkers.

## **Abstracts are arranged according to order of presentation**

### **Modeling and Design of Protein Dynamics and Function**

A. J. Rader, PhD, and M. R. Betancourt, PhD

*Department of Physics and Center for Mathematical Biosciences, IUPUI*

Our research involves numerical studies of protein structure, dynamics, function, and design. Efficient numerical methods and simplified models to represent these proteins are being developed to study complex biophysical phenomena such as protein folding and protein interactions. Investigation of the process of protein folding includes the determination of the protein structure, stability, and folding dynamics for individual and families of proteins. In addition, numerical methods are also being developed for the design of novel proteins with specific physical and functional properties. Physical aspects of protein interactions, especially those that involve large-scale conformational changes or multi-molecular complexes are also studied. One of the main biological motivations behind this work is the ability to deduce the functions of biological molecules from computational analysis of their structures. These functions are inferred from simulations of the dynamics and flexibilities of molecular structures.

### **Dynamics of Basal Ganglia Circuits in Parkinson's Disease**

Leonid Rubchinsky, PhD,<sup>a,b</sup> Choongseok Park, PhD,<sup>a</sup> Alexey Kuznetsov, PhD,<sup>a</sup>  
Robert M. Worth, MD, PhD<sup>a,c</sup>

<sup>a</sup> *Department of Mathematical Sciences and Center for Mathematical Biosciences, IUPUI*

<sup>b</sup> *Stark Neurosciences Research Institute, IU School of Medicine*

<sup>c</sup> *Department of Neurological Surgery, IU School of Medicine*

Parkinson's disease (PD) is one of the major neurodegenerative disorders affecting up to a million Americans. While it is well known that PD results from the degeneration of dopaminergic neurons, little is understood about how this loss results in motor symptoms of PD, despite physiological studies in rodent and primate PD models and in PD patients. We study dynamics of individual neuronal units and neuronal populations (as revealed by the recordings of local field potentials) in Parkinsonian patients. The methods and ideas of nonlinear dynamics are used to get insight in the pathological physiology of basal ganglia motor control circuits of Parkinsonian brain. Some preliminary results of this project will be presented. Understanding of the dynamical nature of activity in affected neuronal circuits will advance us closer to the possibility of efficient control of basal ganglia through adaptive chronic deep brain stimulation.

### **Solving Differential Equations with Multiple Scales: Parallel Computing, Poly-algorithms, Preconditioning and All That**

Raymond C. Y. Chin, PhD,<sup>a</sup> Bruno Bieth<sup>a</sup>, and Lang Li, PhD<sup>b</sup>

<sup>a</sup> *Department of Mathematical Sciences and Center for Mathematical Biosciences, IUPUI*

<sup>b</sup> *Division of Biostatistics, IU School of Medicine*

This is a survey on developing fast and accurate methods for solving differential equations imbued with multiple scales using parallel computing, structure of the solution in the form of preconditioning, and a mixed symbolic and numerical computational methods forming a poly-algorithm. The method developed will be applied to solving parameter estimation problems in the biosciences.

## **Abstracts are arranged according to order of presentation**

### **Statistical Principles and the Role of a Statistician at the Dawn of the Omics Era: Who Should You Look for—A User or a Developer?**

Samiran Ghosh, PhD

*Department of Mathematical Sciences and Center for Mathematical Biosciences, IUPUI*

With the recent advent of computer technology, a new paradigm has begun where complex biological system can be analyzed in a more useful fashion.

In short, this mixing of computational and biological sciences popularly known as “omics” era of biomedical research, often produces “modern data”, which is high-dimensional, noisy and contains a lot of irrelevant predictors. As the new technology shows its promises, it also throws exciting challenges. As it often happens, with the excitement of new technology basic principles of reproducibility, scalability and other design issues are often undermined. More importantly mathematical principles governing many traditionally used statistical techniques are often not satisfied. Hence the routine analysis produces result which is far from being optimal. It is not surprising that we see more and more evolving “tailor-made” statistical/mathematical techniques to be applicable towards a specific bioinformatics study. Over the time underlying statistical principles behind those “tailor-made” techniques are slowly emerging. Robustness, Regularization, Scalability and Adaptive learning are some of those key concepts. Nevertheless methodological innovations are still lagging and are far from being standardized to be used as a “canned package” for data analysis. This is high time we see a pressing need for a new statistical as well as algorithmic development.

The true word behind the success story is “collaboration” between domain experts and quantitative scientists. In this talk I will reiterate some of the issues which are practical as well as relevant for modeling “omics” data set.

### **Biomedical Text Mining: From Data to Information to Knowledge to Hypotheses**

Snehasis Mukhopadhyay, PhD

*Department of Computer & Information Science and Center for Bio-Computing, IUPUI*

A vast volume of biomedical literature (e.g., the PubMed) is currently available in electronic form (data). We will discuss how potentially relevant items (information) can be identified from such large volume of data. We will briefly discuss how relationships among diverse biological objects, e.g., genes, proteins, drugs, diseases, etc (knowledge) can be extracted from such information. Finally, we will present some work on how to extrapolate on existing knowledge to identify new, potentially useful relationships (hypotheses).

Many faculty, researchers, and graduate students have collaborated on grants and papers related to various aspects of the work described. Some of the collaborators include Mathew Palakal (School of Informatics, formerly at Computer & Information Science, IUPUI), Javed Mostafa (Information Science & Biomedical Research & Imaging Center, UNC at Chapel Hill, formerly at SLIS, IUB), Rajeev Raje (Computer & Information Science, IUPUI), Simon Rhodes (IU School of Medicine, formerly at Biology, IUPUI), and David Potter (Biochemistry, IU School of Medicine).

## **Abstracts are arranged according to order of presentation**

### **3D Facial Image Analysis for Medical Diagnosis**

Shiaofen Fang, PhD,<sup>a</sup> Ilona Autti-Rämö, MD, PhD,<sup>b</sup> Åse Fagerlund, PhD,<sup>c</sup>  
Sandra Jacobson, PhD,<sup>d</sup> Luther K. Robinson, MD,<sup>e</sup> H. Eugene Hoyme, MD,<sup>f</sup>  
Sarah N. Mattson, PhD,<sup>g</sup> Edward Riley, PhD,<sup>g</sup> Feng Zhou, PhD,<sup>h</sup> Richard Ward, PhD,<sup>j</sup>  
Elizabeth Moore, PhD,<sup>i,k</sup> Tatiana Foroud, PhD,<sup>h</sup> and the CIFASD\*

<sup>a</sup> Department of Computer & Information Science and Center for Bio-Computing, IUPUI

<sup>b</sup> Department of Child Neurology, HUCH Hospital for Children and Adolescents, Helsinki, Finland

<sup>c</sup> Folkhälsan Research Center, Helsinki, Finland

<sup>d</sup> Department of Psychiatry and Behavioral Neurosciences, Wayne State University

<sup>e</sup> State University of New York at Buffalo School of Medicine

<sup>f</sup> Stanford University School of Medicine

<sup>g</sup> Department of Psychology, San Diego State University

<sup>h</sup> Indiana University School of Medicine

<sup>j</sup> School of Liberal Arts, IUPUI

<sup>k</sup> St. Vincent Hospital, Indianapolis, IN

We will present a new 3D medical image analysis technique for polygon mesh surfaces. This technical is being applied to 3D facial surface analysis for the automatic diagnosis of fetal alcohol syndrome (FAS). The goal is to explore the natural patterns and 3D facial features to provide diagnostic information. Our approach is based on a digital geometry analysis framework that applies machine learning and pattern recognition techniques to polygon mesh surface data from 3D laser scanners and other sources. Novel 3D geometric features are identified, extracted and analyzed to determine the most discriminatory features which can be used for automated diagnosis of FAS. This is a collaborative project within the NIH funded international consortium for FASD (CIFASD).

### **Secure and Seamless Medical Information System**

Xukai Zou, PhD,<sup>a</sup> Yuanshun Dai<sup>b</sup>, PhD, Bradley Doebbeling, MD,<sup>c,d</sup> and John Burke<sup>c</sup>

<sup>a</sup> Department of Computer & Information Science and Center for Bio-Computing, IUPUI

<sup>b</sup> Formerly of Computer & Information Science, IUPUI, currently at University of Tennessee Knoxville

<sup>c</sup> Department of Veterans Affairs Medical Center

<sup>d</sup> Regenstrief Institute and IU School of Medicine

The privacy and security of health information is one of the most critical issues that people are concerned about in adopting and using medical information system (MIS). In addition, data sharing and exchange across hospitals and organizations has become an important yet challenging issue in terms of security, integration, heterogeneity, and scalability. Via the collaboration with the Department of Veterans Affairs (VA) Indianapolis Medical Center and by using our advanced technology which possesses unique security and integration capabilities, the computer science faculty at IUPUI has implemented fundamental security modules and integrated them into VA's VISTA/CPRS/VistaWeb MIS system seamlessly without modifying any existing components. In this presentation, we will introduce such an advanced technology, the powerful features of the integrated system, its fundamental and infrastructural support to clinical practice, scientific biomedical research and tele-medicine, and its applicability to other mission-critical fields. The live demo is also available in the Bio-Computing Center.