EITC-Bio 2008

The First Annual EITC-Bio Workshop
(EITC-Bio 2008)

Synergy of Bioinformatics and Biomedical Research

Friend Center, Princeton University
Princeton, New Jersey, U.S.A.

Saturday, June 7, 2008

Table of Contents

Welcome Message

The world is in the midst of an information and communication technological revolution that is transforming almost every aspect of our lives. The intersection of information technology and biotechnology has become critically important because of the vast amount of data involved in the study of biology. Bioinformatics is very much a discipline in expansion as evidenced by the convergence of Biology, Computer Science, Information and Communication Technology, Mathematics and Statistics.

Bioinformatics highlights the application of statistics, data mining, artificial intelligence, neural networks, machine learning and natural language processing techniques to computationally difficult problems in molecular biology. It is dedicated to provide researchers the knowledge and skills necessary for the invention of algorithms and the creation of computational systems that facilitate the understanding of biological processes and application of these tools and methods to individuals and communities through public health and prevention programs. Meanwhile, advances in high-throughput biotechnology and novel bioassays at the single-cell level have fundamentally changed the way people study biology. Following the decyphering of human genome at the turn of the century, interplay of bioinformatics and novel biotechnology has brought a new revolution in biomedical research.

The EITC-Bio Workshop will focus on the current research and development frontiers in both academia and industry. This year, we will invite leading scientists to present and discuss how the synergy of bioinformatics and biomedical science may further our understandings in biology and medicine, and faciliate our combat against diseases. The full-day event covers latest advances in various topics including computational biology, systems biology, cancer, immunology, and high-throughput screening. We expect our program will engage dialogues across disciplines and invite discussions in the forefront of biomedical informatics. We sincerely invite you to join us for this exciting event.

Li-San Wang, University of Pennsylvania
Yibin Kang, Princeton University

Planning Committee

Program Committee

Shridar Ganesan		UMDNJ	(Cancer Biology)
Frank Hsu	許德標	Fordham University	(Algorithmics and Computational Biology I)
Yibin Kang	康毅滨	Princeton University	(Co-Chair)
Ming OuhYoung	歐陽明	National Taiwan University	(Algorithmics and Computational Biology II)
Li-San Wang	王立三	University of Pennsylvania	(Co-Chair)
Cathy H. Wu	吳慧華	Georgetown University	(Bioinformatics and Systems Biology)

Steering Committee

Eric Y. Chuang	莊曜宇	National Taiwan University
Frank Hsu	許德標	Fordham University
Yibin Kang	康毅滨	Princeton University
Ming OuhYoung	歐陽明	National Taiwan University
Li-San Wang	王立三	University of Pennsylvania (Coordinator)
Cathy H. Wu	吳慧華	Georgetown University Medical Center

Organizing Committee

Kevin T.Y. Hsu	徐子淵	IBM
W.-J. (Adam) Lee	李偉智	University of Maryland (Proceedings)
Yibin Kang	康毅滨	Princeton University
Hwa-Han Wang	王華漢	EITC (Co-Chair)
Wei-Hsing Wang	王維興	EITC (Co-Chair)
Li-San Wang	王立三	University of Pennsylvania
Alvin W.C. Wong	翁唯城	EITC and University of Texas at Dallas (Webmaster)

Conference Program

Time	Schedule	#Speakers
9:00 AM	Keynote	1
9:50 AM	Session 1: Bioinformatics and Systems Biology (I)	3
11:05 AM	*break*	
11:20 AM	Session 2: Cancer Biology (I)	3
12:35 PM	Lunch + Posters	
1:25 PM	Session 3: Bioinformatics and Systems Biology (II)	3
2:40 PM	*break*	
2:55 PM	Session 4: Cancer Biology (II)	2
3:45 PM	*break*	
4:00 PM	Session 5: Bioinformatics and Systems Biology (III)	3
5:15 PM	*break*	
5:30 PM	Panel Discussion	
6:30 PM	*adjourn*	
7:00 PM	Dinner by invitation only	1

Keynote

Speaker: **Jenn-Kang Hwang, Ph.D.**
Dean, College of Biological Science and Technology
National Chiao Tung University (Hsinchu, Taiwan)
On the protein structure-dynamics-function relationship

Session 1: Bioinformatics and Systems Biology (I)

Session Organizer: **Cathy Wu, Ph.D.**
Professor of Biochemistry and Molecular & Cellular Biology
Director of Protein Information Resource
Georgetown University (Washington, DC, USA)

Hai Hu, Ph.D.
Deputy Chief Scientific Officer and Senior Director of Biomedical Informatics
Windber Research Institute (Windber, PA, USA)
Application of Biomedical Informatics to Translational Research

Guang Yao, Ph.D.
Postdoctoral Fellow
Molecular Genetics and Microbiology, Duke University (Durham, NC, USA)
Biological Networks: A bistable Rb–E2F switch underlies the restriction point

Zhang-Zhi Hu, M.D.
Associate Professor and Associate Team Lead
Protein Information Resource
Georgetown University Medical Center (Washington, DC, USA)
Biological Pathway and Network Analysis for Functional Interpretation of Large-Scale Omics Data

Session 2: Cancer Biology (I)

Session Organizer: **Shridar Ganesan, M.D.**
Assistant Professor
University of Medicine and Dentistry of New Jersey (Newark, NJ, USA)

Carlo Maley, Ph.D.
Assistant Professor
The Wistar Institute (Philadelphia, PA, USA)
Solving the Puzzle of Metastasis with an Agent-Based Model

Shridar Ganesan, M.D.
Assistant Professor
University of Medicine and Dentistry of New Jersey (Newark, NJ, USA)
Biology of Basal-like breast cancer

Kim Hirshfield, M.D., Ph.D.
Assistant Professor of Medicine
University of Medicine and Dentistry of New Jersey (Newark, NJ, USA)
Polymorphisms in P53 Pathway Genes and Clinical Associations in Breast Cancer

Session 3: Bioinformatics and Systems Biology (II)

Session Organizer: **Ouhyoung Ming, Ph.D.**
Professor of Computer Science and Information Engineering
Deputy Dean, College of Electrical Engineering and Computer Science
National Taiwan University (Taipei, Taiwan)

Jason T.L. Wang, Ph.D.
Professor, Department of Computer Science
Director, Data and Knowledge Engineering Lab and Bioinformatics Center
New Jersey Institute of Technology (Newark, NJ, USA)
Detecting RNA Motifs with RADAR

Jung-Hsien Chiang, Ph.D.
Professor, Institute of Medical Informatics
National Cheng-Kung University (Tainan, Taiwan)
Data-Driven Computational Function Association Networks in Cancer Study

Spencer Chiang-Ching Huang, Ph.D.
Assistant Professor of Preventive Medicine
Northwestern University (Evanston, IL, USA)
Identification of robust genomic signatures via sequential cross validation

Session 4: Cancer Biology (II)

Session Organizer: **Yibin Kang, Ph.D.**
Assistant Professor
Department of Molecular Biology
Princeton University (Princeton, NJ, USA)

Guohong Hu, Ph.D.
Postdoctoral Fellow, Department of Molecular Biology
Princeton University (Princeton, NJ, USA)
Genomic Gain of 8q22 Activates Metadherin and Promotes Chemoresistant Metastasis of Poor-Prognosis Breast Cancer

Anant Madabhushi, Ph.D.
Assistant Professor, Department of Biomedical Engineering
Rutgers the State University of New Jersey (Piscataway, NJ, USA)
Computer-aided Diagnosis of Breast, Prostate Cancer from Digitized Histopathology

Session 5: Bioinformatics and Systems Biology (III)

Session Organizer: **Frank Hsu, Ph.D.**
Clavius Distinguished Professor of Science and Associate Chair for Graduate Studies
Department of Computer & Information Sciences
Fordham University (Bronx, NY, USA)

Henry Horng-shing Lu, Ph.D.
Professor, Institute of Statistics
National Chiao-Tung University (Hsinchu, Taiwan)
Multidimensional scaling for large genomic data sets

Nickolas Kintos, Ph.D.
Visiting Assistant Professor, Department of Mathematics
Fordham University (Bronx, NY, USA)
Comparing Projection Neuron- and Neuromodulator-Elicited Network Oscillations: A Modeling Study

Mingyao Li, Ph.D.
Assistant Professor, Department of Biostatistics and Epidemiology
University of Pennsylvania (Philadelphia, PA, USA)
Modeling Genetic Inheritance of Copy Number Variations

Dinner Speaker

Cathy Wu, Ph.D.
Professor of Biochemistry and Molecular & Cellular Biology
Director of Protein Information Resource
Georgetown University (Washington, DC, USA)
Integrative Science for the 21st Century: Bioinformatics, Systems Biology and Translational Research

Keynote

On the protein structure-dynamics-function relationship

Jenn-Kang Hwang

Professor, Institute of Bioinformatics
National Chiao Tung University
Hsinchu 300, Taiwan
Republic of China
Tel: +886-3-513-1337, Fax: +886-3-572-9288
Email: jkhwang@faculty.nctu.edu.tw

ABSTRACT

We developed a simple method to compute correlation of fluctuations directly from protein structure. This method, referred to as the protein fixed-point model, is based on the positional vectors of atoms issuing from the fixed point, which is the point of the least fluctuations in proteins. One corollary from this model is that atoms lying on the same shell centered at the fixed point will have the same thermal fluctuations. In practice, this model provides a convenient way to compute the average dynamical properties of proteins directly from the geometrical shapes of proteins without the need of any mechanical models, and hence no trajectory integration or sophisticated matrix operations are needed. As a result, it is more efficient than molecular dynamics simulation or normal mode analysis. Our result shows that the fixed-point model is indeed quite general and will be a useful tool for high throughput analysis of dynamical properties of proteins. In addition, taking advantage of the distinct dynamical properties of active-site residues, we are able to use our approach to detect catalytic residues from protein structures. We believe that our method will be a useful tool for detection of possible active sites from protein structures to complement other existing methods.

BIOGRAPHY

Jenn-Kang Hwang received his B.S. in Agricultural Chemistry from National Taiwan University and his Ph.D. in Physical Chemistry from University of Southern California under Professor Arieh Warshel. He joined the Department of Life Sciences in National Tsing Hua University in 1993. He then moved to National Chiao Tung University in 2000 to help establish the Institute of Bioinformatics, the first one of its kind in Taiwan. Currently, he is Dean of College of Biological Science and Technology in National Chiao Tung University.He is on the editorial boards of Proteins: Structure, Function and Bioinformatics; Recent Patents on Anti-Infective Drug Discovery; the International Journal of Medical Engineering and Informatics. He is Director of the Center for Biomedical and Biological Engineering supported by Ministry of Education. He is currently President of Bioinformatics and Systems Biology Society, Taiwan. Professor Hwang's research interests include protein structure prediction and analysis, protein stability, protein subcellular localization, protein dynamics and molecular simulation.

Session 1: Bioinformatics and Systems Biology (I)

Session Organizer & Chair

Cathy H. Wu

Professor, Department of Biochemistry and Molecular & Cellular Biology
Director, Protein Information Resource
Georgetown University Medical Center
Washington, DC 20007
Tel: 202-687-1039, Fax: 202-687-0057
Email: wuc@georgetown.edu

BIOGRAPHY

Academic/Professional Appointments (current)

- Professor, Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center (GUMC)
- Professor, Department of Oncology, GUMC
- Director, Protein Information Resource, GUMC
- Director, Bioinformatics Track, M.S. Degree in Biochemistry & Molecular Biology, GUMC
- Editorial Board, Journal of Proteomics and Bioinformatics
- Council, Human Proteome Organization (HUPO)
- Board of Directors, US Human Proteome Organization (US HUPO)
- Advisory Board, Protein Data Bank (PDB)
- Advisory Committee, Protein Structure Initiative, NIGMS, National Institutes of Health (NIH)
- TeraGrid User Advisory Committee, National Science Foundation (NSF)
- Education Committee, International Society for Computational Biology (ISCB)
- Advisory Board, Association of Chinese Bioinformaticians

Brief Bio

Cathy H. Wu received her B.S. in Plant Pathology from National Taiwan University in 1978 and Ph.D from Purdue University in 1984. She conducted postdoctoral research in molecular biology, and later obtained a second M.S. degree in Computer Science. With background and experience in both biology and computer science, she has conducted bioinformatics and computational biology research for 20 years. She has developed several protein classification systems and databases, and led the Protein Information Resource (PIR) since 1999. Dr. Wu has served on several advisory boards and many bioinformatics grant review panels at NIH, NSF and DOE. She has served on numerous Program Committees for international bioinformatics and proteomics conferences, such as the International Conference on Intelligent Systems for Molecular Biology (ISMB) and the US HUPO Annual Conference. She has published about 130 peer-reviewed papers and three books, and given more than 100 invited lectures. Her research interests include protein evolution-structure-function relationships, proteomic informatics and computational systems biology, biomedical text mining and ontology, and bioinformatics cyberinfrastructure.

Session 1: Bioinformatics and Systems Biology (I)

Application of Biomedical Informatics to Translational Research

Hai Hu

Deputy Chief Scientific Officer and Sr. Director of Biomedical Informatics
Windber Research Institute, 620 7th St., Windber, PA 15963
Tel: (814)361-6903, Fax: (814)467-6334
Email: h.hu@wriwindber.org

ABSTRACT

This lecture will focus on the application of biomedical informatics to translational research. We define biomedical informatics as the management and usage of biomedical information encompassing clinical informatics, public health informatics, and bioinformatics.This definition is increasingly important in the genomic era, due to the sheer volume of the involved data, as new concepts and technologies enter into medical practice and related basic research. This multidisciplinary subject requires new types of information management and data analysis that relies on sophisticated statistical and computational technologies. Translational research, on the other hand, is often seen as the rapid inclusion of the results of basic biological research into clinical practice, i.e. "bench to bedside".In our practice we have found that it is equally important to have clinical needs feeding into the framing of basic research questions, more of a "bedside – bench – bedside" cycle which further requires a strong biomedical informatics infrastructure. In this lecture, I will address the application of biomedical informatics to translational research from clinical, molecular study, informatics, and data analysis perspectives, using practical examples from a non-profit organization.

Reference: Hai Hu, Richard J. Mural, and Michael N. Liebman (Editors). *Biomedical Informatics: A Translations Approach*. Artech Publishing House.(Publication date: July 2008)

BIOGRAPHY

Hai Hu received his B.S. in Physics from Nanjing Normal University, Nanjing, China, in 1984; M. S. in Speech Signal Processing from the Institute of Acoustics of the Chinese Academy of Sciences, Beijing, China in 1987; and Ph. D. in Biophysics from the State University of New York at Buffalo, USA in 1995. He received a postdoctoral training in Molecular and Cellular Cardiology at the Johns Hopkins University School of Medicine, where his research received fellowships from the National Institute of Health and the American Heart Association. His educational background also includes computer engineering and statistics.

He is currently Deputy Chief Scientific Officer and Senior Director of Biomedical Informatics of the Windber Research Institute with interests in data integration, data analysis, and data mining, as well as algorithm and analytical tool development. Before joining the Institute he was Group Leader and Senior Bioinformatics Scientist at AxCell Biosciences. He has co-edited two books, and published more than 30 peer-reviewed papers and book chapters on subjects including biomedical informatics, bioinformatics, proteomics, and genomics. He serves as a peer-reviewer for a number of journals including Bioinformatics, Proteomics, and Circulation Research. He also serves on the editorial board of two journals. In addition, he is often invited to present at national and international scientific and business conferences, and sometimes serves as bioinformatics program chair or session chair. He holds a joint faculty appointment at the Uniformed Services University of the Health Sciences, and also serves as an Invited Professor (特聘教授) of the Shanghai Center for Bioinformation technology. He is a member of the *International Society for Computational Biology*.

Session 1: Bioinformatics and Systems Biology (I)

A bistable Rb–E2F switch underlies the restriction point

Guang Yao

Postdoctoral Fellow, Department of Molecular Genetics and Microbiology
Duke University
Durham, NC 27708
Tel: 919-684-8709, Fax: 919-681-8973
Email: yao00003@mc.duke.edu

ABSTRACT

The restriction point (R-point) marks the critical event when a mammalian cell commits to proliferation and becomes independent of growth stimulation. It is fundamental to normal differentiation and tissue homeostasis, and appears dysregulated in virtually all cancers. Although the R-point has been linked to various activities involved in regulation of the G1-S transition of the mammalian cell cycle, the underlying mechanism remains elusive. Using single-cell measurements, here we show that the Rb-E2F pathway functions as a bistable switch to convert graded serum inputs into all-or-none E2F responses. Once turned ON by sufficient serum stimulation, E2F can memorize and maintain this ON state, independent of continuous serum stimulation. We further show that, in both critical dose and timing responses, bistable E2F activation directly correlates with a cell's ability to traverse the R-point. This work is a joint research project with T Lee, and Drs. S Mori, JR Nevins, and L You.

BIOGRAPHY

Guang Yao received his B.S. in Molecular Biology from University of Science and Technology of China (USTC) in 1996 and his Ph.D. in Caner Biology from University of Wisconsin at Madison in 2002. He worked as a research associate with his Ph.D. advisor briefly after graduation before moving to Duke University in 2003. He is currently doing his postdoctoral training with Dr. Joseph R. Nevins at Duke University Medical Center.

Guang Yao's current research interest is focused on systems biology with an emphasis of high-resolution single-cell measurements coupled with mathematical modeling, in dissecting biological circuits controlling cell proliferation and programmed cell death.

Session 1: Bioinformatics and Systems Biology (I)

Biological Pathway and Network Analysis for Functional Interpretation of Large-Scale Omics Data

Zhang-Zhi Hu

Research Associate Professor
Protein Information Resource, Department of Biochemistry and Molecular & Cellular Biology
Georgetown University Medical Center, Washington, DC 20007
Phone: (202) 687-1255 Fax: (202) 687-0057
Email: zh9@georgetown.edu

ABSTRACT

The last decade has seen rapid expansion of genomics, transcriptomics, proteomics and other "–omics" as applied to biomedical researches. However, making senses out of the deluge of omics data remains challenging and requires effective bioinformatics approaches for data integration, analyses, and interpretation, which has become an active area of bioinformatics research, and is crucial to the realization of full potentials that omics are hoped to bring. We took a protein-centric data integration and analysis approach and developed iProXpress bioinformatics system for functional analysis of large-scale data, primarily from gene expression and proteomics studies. The system contains three major components: a data warehouse with information derived from over 90 databases, analytical tools for sequence analysis and functional annotation, and a graphical user interface for iterative and comparative function and pathway analysis. The system's unique features include its comprehensiveness of protein sequence coverage and annotation, high protein mapping rate of expression data including protein spliced forms, and its versatility of use on different types of omics data. We applied this approach to several cancer studies for identifying pathways and networks, including organelle biogenesis in melanocytes, signaling and metabolic pathways in hormone- or radiation-resistant breast cancer and other cells. Our studies highlight the need of functional profiling, pathway/network analysis coupled with expert-guided examination for omics data interpretation and for hypothesis formulation.

BIOGRAPHY

Zhang-Zhi Hu received his B.S./M.D in medicine from Wannan Medical College, Anhui China in 1984, and his M.S. in Physiology/Endocrinology from Beijing Medical University (now Beijing University Health Science Center) in 1989.He received NIH Intramural Research Training Award during 1993-1998 to carry out his post-doctoral training in molecular endocrinology at the National Institute of Child Health and Human Development (NICHD). He received the NIH Fellows Award for Research Excellence in 1996 for his findings of multiple and tissue-specific prolactin receptor gene promoters. He then continued his research at the NICHD for three years as a staff fellow before he joined the Protein Information Resource (PIR) at the National Biomedical Research Foundation (NBRF), Washington DC, as a senior bioinformatics scientist. In 2005, he became a Research Associate Professor, Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center, and an Associate Team Lead of Protein Science at the PIR. He is a Guest Professor of Beijing University Health Science Center since 2005.

Zhang-Zhi Hu's current research interests include protein bioinformatics and computational systems biology. His primary research focuses on the understanding of biological pathways and networks through large-scale omics data integration and analyses and associated tools development, as applied to common diseases such as cancers, endocrine and immune disorders. He is also interested in and actively works on biomedical text mining, an emerging area important for database annotation and data integration. His research has been funded by NIH, NSF, and DOD. He has over 45 journal publications and book chapters.

Session 2: Cancer Biology (I)

Session Organizer & Chair

Shridar Ganesan

Assistant Professor,Departments of Medicine and Pharmacology
The Cnncer Institute of New Jersey and the Robert Wood Johnson Medical School, UMDNJ
New Brunswick, NJ 08903
Tel: 732 235 5211, Fax: 732 235 5331
Email: ganesash@umdnj.edu

BIOGRAPHY

Shridar Ganesan received his A.B. in Chemistry from Princeton University in 1985, and his M.D., and Ph.D. from Yale University in 1993. He complete his medical residency at the Brigham and Women's Hospital in Boston, MA, and his medical oncology fellowship at the Dana-Farber Cancer Institute in Boston, MA. He did post-doctoral research in the laboratory of David Livingston and joined the faculty at DFCI and Harvard Medical School in 2000.Shridar Ganesan then joined the faculty at the Cancer Institute of New Jersey and Robert Wood Johnson Medical School, UMDNJ in 2005 where he is currently an Assistant Professor of Medicine and Pharmacology.

Session 2: Cancer Biology (I)

Solving the Puzzle of Metastasis with an Agent-Based Model

Carlo C. Maley

Assistant Professor
Systems Biology Division
Molecular and Cellular Oncogenesis Program
The Wistar Institute
3601 Spruce St., Philadelphia, PA 19104
Tel: 215-495-6838, Fax: 215-495-6829
Email: cmaley@alum.mit.edu

ABSTRACT

Prior to metastasis, most cancers can be treated surgically, with good prognosis. After metastasis, systemic therapy is required and is rarely curative. Thus, metastasis represents one of the most important transitions in neoplastic progression. The evolution of metastasis is a puzzle because a metastatic clone in the primary tumor is at a competitive disadvantage relative to non-metastatic clones and so should be unable to expand to large numbers. The vast majority of cells that emigrate from the primary tumor fail to establish metastases. So a metastatic clone effectively wastes much of its reproductive potential compared to a non-metastatic clone as they compete for space and resources in the primary tumor. We have developed an agent-based model that shows that heterogeneity of resources in both space and time, caused by poorly regulated angiogenesis and hypoxia, selects for increased cell migration within the primary tumor. Cells that are able to locate unutilized resources are better able to proliferate than their sedentary competitors. We hypothesize that this selection for cell migration leads to increased cell emigration from the primary tumor as a side effect. This would explain why primary tumors with high metastatic potential may be detected early in progression. Expression array studies may be detecting the conditions that select for cell migration or may be detecting increased cell migration within the primary tumor. Our results would also explain why hypoxia, particularly transient periods of hypoxia, are associated with increased metastasis. Our results may lead to better assays to measure the risk of metastasis and to interventions that may reduce that risk.

BIOGRAPHY

Carlo C. Maley received his B.A. in computer science and psychology from Oberlin College in 1991, his M.Sc. in zoology from University of Oxford in 1993 and his Ph.D. in computer science from MIT in 1998. He carried out his postdoctoral training with Prof. Stephanie Forrest at the University of New Mexico and then Dr. Brian Reid at the Fred Hutchinson Cancer Research Center. He is currently an assistant professor at the Wistar Institute and a member of the Genomics and Computational Biology as well as the Cellular and Molecular Biology graduate programs at the University of Pennsylvania. He has served on the National Cancer Institute's Translational Research Working Group round tables, the AACR Task Force for Cancer Prevention and has been an advisor to the National Commission on Digestive Diseases.

Carlo Maley's current research interests include applying evolutionary theory to the understanding of neoplastic progression, therapeutic resistance, and the development of cancer prevention strategies. His research activities have been supported by the National Institutes of Health, Department of Defense, the Sloan Foundation, the Pew Charitable Trust, the PhRMA Foundation, the McLean Contributionship and the Pennsylvania Department of Health. He is the recipient of the first Landon AACR Innovator Award for Cancer Prevention Research. He has authored and coauthored a mere 24 reviewed scientific papers as well as 14 chapters and reviews.

Session 2: Cancer Biology (I)

Biology of Basal-like Breast Cancer

Shridar Ganesan

Assistant Professor,Departments of Medicine and Pharmacology
The Cnncer Institute of New Jersey and the Robert Wood Johnson Medical School, UMDNJ
New Brunswick, NJ 08903
Tel: 732 235 5211, Fax: 732 235 5331
Email: ganesash@umdnj.edu

ABSTRACT

Gene expression analysis has identified basal-like breast cancer (BLC) as an biologically distinct subclass of human breast cancer. BLC account for 15% of human breast cancer. These tumors tend to occur in younger women, have no targeted therapy and have a poor overall prognosis. The breast cancers that arise in women with BRCA1 mutations mostly cluster with BLC by gene experession. SNP-array based analysis has shown that sporadic BLC have highly similar pattern of genomic changes as that seen in BRCA1 -/- tumors, consistent with the presence of similar defects in genomic stability. This observation suggests that sporadic BLC may have abnormalities in BRCA1-related DNA repair pathways.. As the repair defect in BRCA1 -/- cells render them sensitive to certain classes of DNA damaging agents, such as cis-platin, this observation suggests that sporadic BLC may also be vulnerable to these agents. A search for specific abnormalities in DNA repair proteins in BLC has identified that a subset ofsporadic BLC have abnormal expression of 53BP1, a protein that has a critical role in the DNA repair and checkpoint control pathways. These observations are consistent with sporadic BLC having profound defects in DNA-repair and suggest specific therapeutic approaches that exploit these vulnerabilities.

BIOGRAPHY

Shridar Ganesan received his A.B. in Chemistry from Princeton University in 1985, and his M.D., and Ph.D. from Yale University in 1993. He complete his medical residency at the Brigham and Women's Hospital in Boston, MA, and his medical oncology fellowship at the Dana-Farber Cancer Institute in Boston, MA. He did post-doctoral research in the laboratory of David Livingston and joined the faculty at DFCI and Harvard Medical School in 2000.Shridar Ganesan then joined the faculty at the Cancer Institute of New Jersey and Robert Wood Johnson Medical School, UMDNJ in 2005 where he is currently an Assistant Professor of Medicine and Pharmacology.

Session 2: Cancer Biology (I)

Polymorphisms in P53 Pathway Genes and Clinical Associations in Breast Cancer

Kim M. Hirshfield

Assistant Professor, Department of Medicine
UMDNJ/Robert Wood Johnson Medical School
The Cancer Institute of New Jersey
New Brunswick, NJ 08901
Tel: 732-235-6028, Fax: 732-235-5331
Email: hirshfie@umdnj.edu

ABSTRACT

Mutations in the tumor suppressor gene p53 are seen in many tumor types, including breast cancer, and may negatively affect response to cancer therapies. Complex regulation of p53 has evolved since heightened activity leads to inappropriate cell death while not enough predisposes to tumorigenesis. The main negative regulator, mdm2, is also dysregulated in many cancers resulting in inefficient p53 response while the structurally-similar mdm4 also interferes with p53 activity. P53-responsive downstream genes, e.g. perp, also play a role in efficiency of the p53 pathway. We have identified single nucleotide polymorphisms in key regulatory genes of the p53 pathway that alter either gene expression or protein function.We have evaluated the role of these gene variants in the age of diagnosis of breast cancer, role in specific molecular and histological subtypes of breast cancer, and recurrence of breast cancer. Gene variants associated with hormone receptor negative breast cancers include p53 and mdm4 while hormone receptor positive correlations include mdm2 and perp. While mdm2 phenotypes associate with estrogen receptor positivity, perp has its strongest associations in progesterone receptor positive breast cancers. This is thought to be related to its role in the desmosomes and cell adhesion. Furthermore, the perp variant highly correlates with risk of recurrence of breast cancer and the rapidity with recurrence. Its aggressive nature parallels hormone receptor negative breast cancers. This work is performed in collaboration with Dr. Arnold Levine and members of the Institute for Advanced Studies.

BIOGRAPHY

Kim Hirshfield received her B.S. in Biochemistry from Rutgers University in 1988 and her Ph.D. Biology with distinctions in Biochemistry from Johns Hopkins University in 1994. She received her M.D. from UMDNJ/Robert Wood Johnson Medical School where she also completed her residency in Internal Medicine and fellowship in Medical Oncology. She carried out her postdoctoral training with Dr. Arnold Levine at The Cancer Institute of New Jersey where she was appointed an Assistant Professor. Her clinical expertise is in early stage breast cancer and high risk breast abnormalities. She is affiliated with the Department of Molecular Genetics, Microbiology, and Immunology at UMDNJ/Robert Wood Johnson Medical School and the Division of Cancer Genomics at The Cancer Institute of New Jersey. She serves as the Principal Investigator for a tissue and blood banking protocol and breast cancer database at The Cancer Institute of New Jersey.

Kim Hirshfield's current research interests include identifying functional single nucleotide polymorphisms in genes important in breast cancer development and elucidating the molecular mechanisms for their functionality. She is especially interested in identifying women at risk for early age of diagnosis and for those with poor prognosis based on their genetic signature. She is also applying her findings to other tumor types. She has been invited to give an educational session for AACR on the use of single nucleotide polymorphisms as prognostic indicators in cancer outcomes. Her research activities have been supported by the New Jersey Commission for Cancer Research, The Breast Cancer Research Foundation, National Institutes of Health, and Department of Defense.

Session 3: Bioinformatics and Systems Biology (II)

Session Organizer

Ming Ouhyoung

Professor, Computer Science and Information Engineering
Deputy Dean, College of Electrical Engineering and Computer Science
National Taiwan University
Taipei, 106, Taiwan
Tel: +886-2-2362-5336 ext. 421, Fax: +886-2-2362-8167
Email: mingouhyoung@gmail.com

BIOGRAPHY

Ming Ouhyoung received the BS and MS degree in electrical engineering from the National Taiwan University, Taipei, in 1981 and 1985, respectively. He received the Ph.D degree in computer science from the university of North Carolina at Chapel Hill in Jan., 1990. He was a member of the technical staff at AT&T Bell Laboratories, Middle-town, during 1990 and 1991. Since August 1991, he has been an associate professor in the department of Computer Science and Information Engineering, National Taiwan University. Then since August 1995, he became a professor. He was the Director of the Center of Excellence for Research in Computer Systems, College of Engineering, from August 1998 to July 2000, and was the Chairman of the Dept. of CSIE from August 2000 to July 2002. He was the deputy dean of College of EECS, 2005-2007. He has published over 100 technical papers on computer graphics, virtual reality, and multimedia systems. He is a senior member of ACM and member of IEEE.

Session 3: Bioinformatics and Systems Biology (II)

Chair

Li-San Wang

Assistant Professor of Pathology and Laboratory Medicine
Institute on Aging / Penn Center for Bioinformatics
University of Pennsylvania
Philadelphia, PA 19104
Email: lswang@mail.med.upenn.edu

BIOGRAPHY

Li-San Wang received his B.S. (1994) and M.S. (1996) in Electrical Engineering from the National Taiwan University.He received his M.S. (2000) and Ph.D. (2003) from the University of Texas at Austin, both in Computer Sciences, and was a postdoctoral fellow at the University of Pennsylvania between 2003 and 2006.Currently he is an Assistant Professor of Pathology and Laboratory Medicine and a fellow of the Institute on Aging, University of Pennsylvania. Dr. Wang's research interests include phylogenetics, comparative genomics, and microarray analysis.He has authored twenty six peer-reviewed book chapters and journals on computational biology and bioinformatics, and served on the program and organizing committees of several international workshops and conferences.

Session 3: Bioinformatics and Systems Biology (II)

Detecting RNA Motifs with RADAR

Jason T. L. Wang

Professor, Bioinformatics Program and Computer Science Department
New Jersey Institute of Technology
University Heights, Newark, NJ 07102
Tel: 973-596-3396, Fax: 973-596-5777
Email: wangj@njit.edu

ABSTRACT

RADAR is a software toolkit that provides a multitude of functionality for RNA data analysis and research. It can align structure-annotated RNA sequences so that both sequence and structure information are taken into consideration during the alignment process. This toolkit is capable of performing pairwise structure alignment, multiple structure alignment, database search and clustering. In addition, RADAR provides two salient features: (i) constrained alignment of RNA secondary structures, and (ii) prediction of the consensus structure for a set of RNA sequences. RADAR is able to assist scientists in performing many important RNA mining operations, including the understanding of the functionality of RNA sequences, the detection of RNA structural motifs and the clustering of RNA molecules, among others. This talk focuses on application of RADAR to motif detection in the mRNAs of various organisms, and sketches some of the algorithms employed by RADAR.

BIOGRAPHY

Jason T. L. Wang (http://web.njit.edu/~wangj) received his B.S. in Mathematics from National Taiwan University in 1980 and his Ph.D. in Computer Science from the Courant Institute of Mathematical Sciences at New York University in 1991. Currently he is a Professor of Bioinformatics and Computer Science in the New Jersey Institute of Technology, New Jersey's Science and Technology University, and directs the University's Data and Knowledge Engineering Laboratory. He is the executive editor of World Scientific Book Series on Science, Engineering, and Biology Informatics (SEBI) and serves on the editorial advisory boards of 9 journals including International Journal of Data Mining and Bioinformatics, International Journal of Soft Computing and Bioinformatics, International Journal of Computational Intelligence in Bioinformatics and Systems Biology, International Journal of Computational Biology and Drug Design, Information Systems, Knowledge and Information Systems, International Journal of Information Quality, Intelligent Data Analysis, and The Open Cybernetics and Systemics Journal. In addition, he has served on the program committees of over 100 national and international conferences.

Jason Wang's current research interests include data mining, bioinformatics and systems biology with an emphasis of RNA structure analysis and biological network mining. His research activities have been supported by Novartis Pharmaceuticals Corporation, AT&T Foundation, Alfred P. Sloan Foundation, James S. McDonnell Foundation, Howard Hughes Medical Institute, Department of Defense, National Institutes of Health, and National Science Foundation. He has authored and coauthored over 140 reviewed scientific papers, and published 5 books including Pattern Discovery in Biomolecular Data: Tools, Techniques and Applications (1999, Oxford University Press), Mining the World Wide Web: An Information Search Approach (2001, Kluwer), Computational Biology and Genome Informatics (2003, World Scientific), Data Mining in Bioinformatics (2005, Springer), and Analysis of Biological Data: A Soft Computing Approach (2007, World Scientific).

Session 3: Bioinformatics and Systems Biology (II)

Data-Driven Computational Function Association Networks in Cancer Study

Jung-Hsien Chiang

Professor, Department of Computer Science
National Cheng Kung University
Tainan, 701, Taiwan
Tel: +886-6-2757575 ext 62534, Fax: +886-6-2747076
Email: jchiang@mail.ncku.edu.tw

ABSTRACT

We have implemented a computational analysis tool to construct the function association networks to analyze the enriched gene sets of microarray data in cancer and stem cell. The main goal of the functional network analysis is to study the most significant biological pathways and functions among the differentially expressed genes between leukemia stem cell and normal stem cell. The tool helps scientists to understand the major biological and functional groups, which have been involved in our focused gene list. In the past, the functional groups are displayed in table formats, and the functional group association networks allow users to be able to visualize the relationship among different functional groups and gene annotations. In addition to functional groups association networks, we also implement another algorithm to construct gene-function relationships and to identify important genes with similar functional annotations and these differentially expressed genes are the potential biomarkers for cancers and diseases. We generated a gene list of differentially expressed genes among leukemia stem cell and normal stem cell. From the gene list, we are able to construct functional groups association network and gene function association networks.

BIOGRAPHY

Jung-Hsien Chiang received the B.S. degree in electrical engineering from National Taiwan University of Science and Technology, Taiwan and the M.S. and Ph.D. degrees in computer engineering from the University of Missouri, Columbia, in 1991 and 1995, respectively. Currently, he is Director of Sun® Center of Excellence on Bioinformatics at Taiwan and Professor of the Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan. From 1995 to 1996, he worked as a researcher at the Computer and Communication Laboratory, Industrial Technology Research Institute (ITRI), the largest information technology research institute in Taiwan. His research interests include bioinformatics, systems biology, text mining, fuzzy modeling, and pattern clustering problems. His current contact address is Institute for Systems Biology, Seattle, WA 98103, USA.

Session 3: Bioinformatics and Systems Biology (II)

Identification of Robust Genomic Signatures via Multiple Cross Validation

Chiang-Ching Spencer Huang

Assistant Professor, Department of Preventive Medicine
Northwestern University
Chicago, IL 60611
Tel: 312-503-4833, Fax: 312-908-9588
Email: huangcc@northwestern.edu

ABSTRACT

Microarrays have demonstrated their utility in biomarker research to improve diagnosis and prognosis. A wealth of research has been devoted to identify differentially expressed genes and develop classification models. However, genes identified may not provide insight into biologically significant pathways, raising concern regarding the robustness of genomic signatures from microarrays. This hurdle is largely due to the limited number of samples under study, difficulty in disease ascertainment, and disease heterogeneity. We have developed a multiple cross validation scheme to identify a subset of homogenous samples that are otherwise not detected by unsupervised approaches such as cluster analysis. This subset of homogenous samples can then be used to identify robustly differentially expressed genes. Our method is applied to a gene expression study to identify genes associated with atherosclerosis. Genes identified therein not only exhibit a significantly reduced false discovery rate but also are enriched in several atherosclerosis-related signaling pathways.

BIOGRAPHY

Chiang-Ching Spencer Huang received his B.S. in Applied Mathematics from National Chiao-Tung University in 1987 and his M.S. in Statistics from University of Iowa in 1998 and Ph.D. in Biostatistics from University of Michigan in 2003. He joined the Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University in 2003. Other than his faculty appointment as an assistant professor, he is currently the Director of Public Health Informatics, Northwestern University Biomedical Informatics Center under the newly established Northwestern University Clinical and Translational Sciences Institute.

Chiang-Ching Spencer Huang's current research interests include identification of prognostic and/or diagnostic biomarkers in cancer and cardiovascular diseases, functional-genomic epidemiology, and nutritional metabolomics, especially in the large-scale population studies. His research activities have been largely supported by National Institutes of Health. He has authored and coauthored over 20 reviewed scientific papers.He is a member of American Statistics Association and American Heart Association.

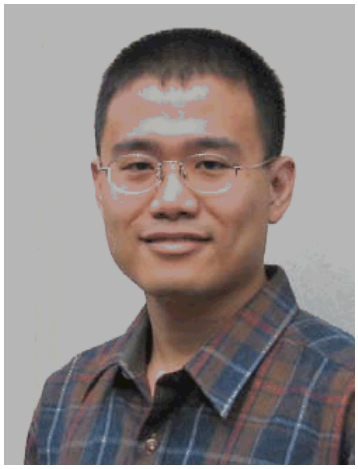
Session 4: Cancer Biology (II)

Session Organizer & Chair

Yibin Kang

Assistant Professor, Department of Molecular Biology
Princeton University
Princeton, NJ 08544
Tel: 669-258-8834, Fax:609-258-2340
Email: ykang@princeton.edu

BIOGRAPHY



Dr. Kang is a native of Fujian Province in Southern China.He was selected to an experimental science class in Beijing University High School at the age of 15 after wining the National Chemistry Competition in 1988. He received his bachelor's degree from the Department of Genetics at Fudan University in Shanghai in 1995.As a graduate student at Duke University, Dr. Kang studied the mechanism of retroviral gene regulation and cellular mRNA export with renowned virologist Bryan Cullen. After completing his graduate study at Duke in just four years and with 11 publications, Dr. Kang joined the Memorial Sloan-Kettering Cancer Center as a postdoctoral fellow with Dr. Joan Massagué in 2000.He conducted ground-breaking research on TGFß signal transduction and functional genomic analysis of breast cancer tissue-specific metastasis. During Dr. Kang's pre- and post-doctoral research career, he published over 20 original articles in leading journals such as Cell, Molecular Cell, Cancer Cell and Genes & Development.

Dr. Kang joined the faculty of Princeton University as an Assistant Professor of Molecular Biology in the fall of 2004. Dr. Kang's research focuses on the molecular mechanisms of breast cancer metastasis, which is responsible for the large majority of cancer deaths. Dr. Kang's laboratory applies a multidisciplinary approach to analyze the molecular basis of cancer metastasis, combining molecular biology and genomics tools with animal models and advanced in vivo imaging technologies.

Dr. Kang's exceptional achievements have been recognized by many prestigious awards, including an American Cancer Society Scholar Award. He was one of the five recipients of the 2006 Department of Defense Era of Hope Scholar Award, intended for exceptionally talented, early-career scientists who have demonstrated that they are the best and brightest in their field through exceptional creativity, vision, and productivity.

Session 4: Cancer Biology (II)

Genomic Gain of 8q22 Activates Metadherin and Promotes Chemoresistant Metastasis of Poor-Prognosis Breast Cancer

Guohong Hu

Postdoctoral Researcher, Department of Molecular Biology
Princeton University
LTL 254, Washington Road, Princeton, NJ, 08544
Tel: 609-258-5939, Fax: 609-258-2340
Email: guohongh@princeton.edu

ABSTRACT

Genetic changes essential for cancer metastasis are often hidden among numerous alterations in the genome of advanced-stage tumors. Here an integrative strategy was used to identify genomic alterations and the target genes that are clinically relevant and functionally significant for breast cancer progression. We designed and applied a computational algorithm to map minimal recurrent genomic alterations associated with poor-prognosis breast cancer.8q22 genomic gain was identified in multiple breast cancer datasets by this approach and validated in an extensive collection of tumor samples. 8q22 amplification leads to elevated expression of the metastasis gene Metadherin (MTDH), which is overexpressed in more than 40% of breast cancers and is associated with poor clinical outcomes.Functional characterization of MTDH revealed its dual role in promoting metastatic seeding and enhancing chemoresistance.Overexpression and knockdown of MTDH elevated and repressed, respectively, the metastasis of breast cancer cells to various organs in animals. Inhibition of MTDH significantly sensitized the cancer cells to anti-neoplastic agents both in vitro and in vivo.Expression profiling and functional analysis further identified ALDH3A1 and MET as two of the downstream genes that mediate the broad spectrum chemoresistance function of MTDH. These findings illustrate the synergistic value of integrating clinical and experimental metastasis research and establish MTDH as an important therapeutic target for simultaneously enhancing chemotherapy efficacy and reducing metastasis risk.

BIOGRAPHY

Guohong Hu received his B.S. in Biology from Tsinghua University in 2000, his M.S in Biology from Rutgers University in 2002 and his Ph.D in Molecular Genetics from Robert Wood Johnson Medical School in 2005. He is currently a postdoctoral research follow in Dr. Yibin Kang's laboratory at the Department of Molecular Biology of Princeton University.

In his graduate school, Guohong Hu's research focused on the development of sensitive gene expression and genotyping microarray systems for cancer profiling and detection, as well as the computational algorithm development for multiplex PCR primer designing and microarray data analysis. Currently he is working with Dr. Kang to study cancer metastasis with an integrative approach combing bioinformatics, animal model and clinical studies.

Session 4: Cancer Biology (II)

Computer-aided Diagnosis of Breast, Prostate Cancer from Digitized Histopathology

Anant Madabhushi

Assistant Professor, Biomedical Engineering Department
Rutgers The State University of New Jersey
Piscataway, NJ 08854
Tel: 732-445-4500, Fax: 732-445-3753
Email: anantm@rci.rutgers.edu

ABSTRACT

Automated detection and segmentation of nuclear and glandular structures is critical for classification and grading of prostate and breast cancer histopathology. In this talk, I will discuss efforts in my lab, Laborator for Computational Imaging and Bioinformatics, work on automated detection and segmentation of structures of interest in digitized histopathology images. The scheme integrates image information from across three different scales: (1) low-level information based on pixel values, (2) high-level information based on relationships between pixels for object detection, and (3) domain-specific information based on relationships between histological structures. Low-level information is utilized by a Bayesian classifier to generate a likelihood that each pixel belongs to an object of interest. High-level information is extracted in two ways: (i) by a level-set algorithm, where a contour is evolved in the likelihood scenes generated by the Bayesian classifier to identify object boundaries, and (ii) by a template matching algorithm, where shape models are used to identify glands and nuclei from the low-level likelihood scenes. Structural constraints are imposed via domain specific knowledge in order to verify whether the detected objects do indeed belong to structures of interest. The utility of the glandular and nuclear segmentation algorithm is demonstrated in accurate extraction of various morphological and nuclear features for automated grading of (a) prostate cancer, (b) breast cancer, and (c) distinguishing between cancerous and benign breast histology specimens. The efficacy of our segmentation algorithm is evaluated by comparing breast and prostate cancer grading and benign vs. cancer discrimination accuracies with corresponding accuracies obtained via manual detection and segmentation of glands and nuclei.

BIOGRAPHY

Dr. Anant Madabhushi received his Bachelors Degree in Biomedical Engineering from Mumbai University, India in 1998 and his Masters in Biomedical Engineering from the University of Texas, Austin in 2000. In 2004 he obtained his PhD in Bioengineering from the University of Pennsylvania. He joined the Department of Biomedical Engineering, Rutgers University as an Assistant Professor in 2005. He is also a member of the Cancer Institute of New Jersey and an Adjunct Assistant Professor of Radiology at the Robert Wood Johnson Medical Center, NJ.

Dr. Madabhushi has nearly 50 publications and book chapters in leading International journals and peer-reviewed conferences. His research interests are in the area of medical image analysis, computer-aided diagnosis, machine learning, and computer vision and in applying these techniques for early detection and diagnosis of prostate and breast cancer from high resolution MRI, MR spectroscopy, protein- and gene-expression studies and digitized tissue histopathology. He is also the recipient of a number of awards for both research as well as teaching, including the Busch Biomedical Award (2006), the Technology Commercialization Award (2006), the Coulter Early Career award (2006), the Excellence in Teaching Award (2007, 2008), the Cancer Institute of New Jersey New Investigator Award (2007), and the Society for Imaging Informatics in Medicine (SIIM) New Investigator award (2008). In addition his research work has also received grant funding from the National Cancer Institute, New Jersey Commission on Cancer Research, and the Department of Defense.

Session 5: Bioinformatics and Systems Biology (III)

Session Organizer & Chair

D. Frank Hsu

Clavius Professor of Science, and
Associate Dean of Graduate School of Arts and Sciences
Fordham University, New York, NY 10458, USA
Tel: 718-817-4483, 212-636-7925
Email: hsu@trill.cis.fordham.edu

BIOGRAPHY

D. Frank Hsu, Ph.D. is Clavius Professor of Science and Professor of Computer and Information Science at Fordham University.He has served as chairman of the Department of Computer and Information Science and now is Associate Dean of the Graduate School of Arts and Sciences at Fordham.Dr. Hsu received a B.S. from National Cheng Kung University (NCKU), Tainan, Taiwan, an M.S. from the University of Texas at El Paso (UTEP), Texas, and a Ph.D. from the University of Michigan. He has held visiting positions at M.I.T., Boston University, Keio University (as IBM Chair Professor), JAIST (as Komatsu Chair Professor), Taiwan University, and the University of Paris-Sud.

Dr. Hsu's research interests are combinatorics, algorithms, and optimizations; network interconnections and communications; informatics and intelligent systems; and information and telecommunication infrastructure.His recent work on informatics and data mining has applications in biomedicine, virtual screening and drug discovery, target recognition and tracking, and information and music retrieval. He has recently developed Combinatorial Fusion Analysis and studied its application to information fusion and knowledge discovery at both the data and decision level.

Dr. Hsu has served on several editorial boards including as Associate Editor of Pattern Recognition Letters and as Editor in Chief (2000-06) and for special issue of the Journal of Interconnection Networks (JOIN). He is a Foundation Fellow of ICA, a senior member of IEEE Computer Society, and a Fellow of the New York Academy of Sciences.

Session 5: Bioinformatics and Systems Biology (III)

Multidimensional scaling for large genomic data sets

Jengnan Tzeng[1], **Henry Horng-Shing Lu**[2], and Wen-Hsiung Li[1,3]

[1]Genomics Research Center, Academia Sinica, Taipei, 115, Taiwan
[2]Institute of Statistics, National Chiao Tung University, Hsinchu, 300, Taiwan
Tel: 886-3-5731870, Fax: 886-3-5728745, Email: hslu@stat.nctu.edu.tw
[3]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

ABSTRACT

Background
Multi-dimensional scaling (MDS) is aimed to represent high dimensional data in a low dimensional space with preservation of the similarities between data points. This reduction in dimensionality is crucial for analyzing and revealing the genuine structure hidden in the data. For noisy data, dimension reduction can effectively reduce the effect of noise on the embedded structure. For large data set, dimension reduction can effectively reduce information retrieval complexity. Thus, MDS techniques are used in many applications of data mining and gene network research. However, although there have been a number of studies that applied MDS techniques to genomics research, the number of analyzed data points was restricted by the high computational complexity of MDS. In general, a non-metric MDS method is faster than a metric MDS, but it does not preserve the true relationships. The computational complexity of most metric MDS methods is over $O(N^2)$, so that it is difficult to process a data set of a large number of genes N, such as in the case of whole genome microarray data.

Results
We developed a new rapid metric MDS method with a low computational complexity, making metric MDS applicable for large data sets. Computer simulation showed that the new method of split-and-combine MDS (SC-MDS) is fast, accurate and efficient. Our empirical studies using microarray data on the yeast cell cycle showed that the performance of K-means in the reduced dimensional space is similar to or slightly better than that of K-means in the original space, but about three times faster to obtain the clustering results. Our clustering results using SC-MDS are more stable than those in the original space. Hence, the proposed SC-MDS is useful for analyzing whole genome data.

Conclusions
Our new method reduces the computational complexity from $O(N^3)$ to $O(N)$ when the dimension of the feature space is far less than the number of genes N, and it successfully reconstructs the low dimensional representation as does the classical MDS. Its performance depends on the grouping method and the minimal number of the intersection points between groups. Feasible methods for grouping methods are suggested; each group must contain both neighboring and far apart data points. Our method can represent high dimensional large data set in a low dimensional space not only efficiently but also effectively.

BIOGRAPHY

Henry Horng-Shing Lu received his Ph.D. and M.S. degrees in Statistics from Cornell University, NY, USA, in 1994 and 1990, respectively, and his B.S. degree in electric engineering from National Taiwan University, Taiwan, ROC, in 1986. He is a Professor in the Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan, ROC. He has been a visiting scholar at UCLA, Harvard University and University of Chicago. His research interests include statistics, medical images, and bioinformatics. He and collaborators have more than 30 journal papers published or accepted, including Journal of the American Statistical Association, Journal of Multivariate Analysis, Statistica Sinica,

Journal of Computational and Graphical Statistics, IEEE Transactions on Reliability/Image Processing/Medical Imaging, Pattern Recognition, Ultrasound in Medicine and Biology, Trends in Genetics, Proceedings of the National Academy of Sciences of the United States of America, Journal of Computational Biology, Bioinformatics, BMC Bioinformatics and so forth.

Session 5: Bioinformatics and Systems Biology (III)

Comparing Projection Neuron- and Neuromodulator-Elicited Network Oscillations: A Modeling Study

Nickolas Kintos

Visiting Assistant Professor, Department of Mathematics
Fordham University
New York, NY 10023
Tel: 212-636-6333, Fax: 212-636-7217
Email: kintos@fordham.edu

ABSTRACT

Neural network activity is often shaped by input from modulatory projection neurons. Although bath application of neuronally-released modulators does not generally mimic network activity elicited by projection neurons, there are some cases where the activity is the same. We examine this issue in the gastric mill (chewing) motor network in the stomatogastric ganglion (STG) of the crab, *Cancer borealis*. Stimulation of the projection neuron MCN1 elicits a gastric mill rhythm (GMR) *in vitro*. The STG terminals of MCN1 have complex synaptic interactions with the gastric mill network during this GMR. Interestingly, without MCN1 participation, bath application of the neuropeptide pyrokinin (PK) elicits a similar GMR. PK is not released by MCN1, and the mechanism that underlies the PK-GMR is unknown.

We use low-dimensional mathematical and higher-dimensional biophysically-realistic models in parallel to propose potential mechanisms by which PK can elicit an MCN1-like GMR. We show that PK can elicit this GMR by activating voltage-gated ionic currents in a specific gastric mill neuron. Our results illustrate how the synaptic actions of a projection neuron can be mimicked by bath application of a neuromodulator through mathematically similar but physiologically distinct mechanisms.

We also compare how the MCN1- and PK-elicited GMRs are coordinated. We show that presynaptic inhibition of MCN1 terminals is necessary for coordinating the MCN1-GMR. In contrast, the PK-GMR can be coordinated by a PK-strengthened gastric mill synapse that is not functional during the MCN1-GMR. We therefore illustrate that distinct mechanisms can regulate the MCN1- and PK-elicited GMRs. (Joint work with Professors Farzan Nadim - NJIT; Michael Nusbaum - U Penn School of Medicine).

BIOGRAPHY

Nickolas Kintos received his B.A. in Mathematics and B.A. in Chemistry from Rutgers University (Newark, NJ) in 2000. He received his M.S. in Applied Mathematics from the New Jersey Institute of Technology (NJIT – Newark, NJ) in 2005. He received his Ph.D. in Mathematical Sciences from NJIT in 2007. Currently he is a Peter M. Curran Visiting Assistant Professor in the Department of Mathematics at Fordham University (New York, NY).

Nickolas Kintos' research interests include singularly perturbed dynamical systems and computational neuroscience. In particular, he builds mathematical models, based upon the experimental data of collaborators, to investigate the dynamics that underlie the activity of neural networks. He is currently a member of the Society for Neuroscience (SfN) and the Society for Industrial and Applied Mathematics (SIAM). He has recently co-authored a research paper (Journal of Computational Neuroscience) and several abstracts for conference proceedings (Society for Neuroscience Meeting, SIAM Conference on the Life Sciences, and East Coast Nerve Net – Marine Biological Laboratories, Woods Hole, MA).

Session 5: Bioinformatics and Systems Biology (III)

Modeling Genetic Inheritance of Copy Number Variations

Mingyao Li

Assistant Professor, Department of Biostatistics and Epidemiology
University of Pennsylvania School of Public Health
Philadelphia, PA 19104
Tel: 215-746-3916, Fax: 215-573-4865
Email: mingyao@mail.med.upenn.edu

ABSTRACT

The analysis of copy number variation (CNV) is now a common approach to identifying disease-causing loci. It has been proposed that these structural variants could also serve as genetic polymorphisms in gene-mapping studies. However, despite increased recognition of their importance, little is known on how to use them in genetic linkage and association analysis. The difficulty lies in the fact that all available CNV detection algorithms are limited to detecting total copy numbers, rather than CNV genotypes, i.e., copy number in each of the two homologous chromosomes. To use CNVs as genetic markers, it is necessary to determine the location of CNVs, the total copy numbers, and the CNV genotypes.

In this talk, I will describe a unified statistical model that addresses all above-mentioned issues. By treating the parents-offspring trio as a unit, our method calls the CNVs of the trio simultaneously, thus avoiding generating calls that are Mendelian inconsistent while maintaining the ability to detect *de novo* CNVs. Extensive simulations indicate that the proposed method significantly increases call rate for inherited CNVs as compared to existing methods. We also applied our method to 28 HapMap trios genotyped using Illumina's 550K SNP array, and detected 2,449 CNVs. Compared to methods that do not use family relationship or use it *a posteriori*, our method detected 25.3% and 7.1% more CNVs, respectively. Our method is general and flexible, and can be applied to data generated from different technical platforms. Moreover, it builds a solid foundation for the development of linkage and association tests on CNVs.

BIOGRAPHY

Mingyao Li received her B.S. in Mathematics from Nankai University, and Ph.D. in Biostatistics from the University of Michigan. In 2006, she joined the Department of Biostatistics and Epidemiology at the University of Pennsylvania as an Assistant Professor.

Her main research area is statistical genetics. In particular, she is interested in developing statistical methods and computational tools for identifying and characterizing genetic variants that influence susceptibility to complex diseases. Much of her research has focused on the use of linkage disequilibrium in the mapping of complex disease susceptibility genes. Her current research work involves evaluating and improving power for genome-wide association studies, and analysis of copy number variations.

In addition to methods development, Dr. Li is also interested in collaborating with researchers seeking to identify complex disease susceptibility genes. Her collaborative research includes studies of the genetics of coronary artery disease, heart failure, childhood obesity, type 2 diabetes, and autism.

Poster Session

Integrative Analysis of Transcriptional Regulation in Growth Hormone Induced Insulin Resistance

Yili Chen

Postdoctoral Researcher, Delaware Biotechnology Institute
University of Delaware
15 Innovation Way, Newark, DE 19711
Tel: 302-831-6168
Email: ychen@udel.edu

ABSTRACT

Growth hormone (GH) is a diabetogenic hormone which induces insulin resistance. GH impairs basal and insulin-stimulated glucose utilization in cultured 3T3-F442A adipocytes, metabolic changes that are consistent with insulin resistance. Microarray analysis of GH-treated 3T3-F442A adipocytes identified 561 probe sets with time-dependent patterns of GH-regulated expression. Deep data mining of this microarray dataset and computational analysis of GH-dependent gene regulation were conducted to predict novel features/factors in GH-regulated gene transcription related to GH induced insulin resistance. An integrative strategy involving a series of computational approaches was applied to hunt for the gene candidates in this biological context. Gene Set Enrichment Analysis (GSEA), to identify GH-regulated pathways relevant to insulin resistance, showed that within the insulin signaling pathway, the GH-regulated genes that encode the negative regulator Suppressor of Cytokine Signaling 2 (Socs2) and the signaling molecule Phosphatidylinositol 3-Kinase Regulatory Subunit p85alpha (Pik3r1) were up-regulated. Real-time PCR confirmed their stimulation by GH at appropriate times. A novel CRC clustering algorithm, applied to identify genes co-regulated by GH, identified not only Socs2 and Pik3r1, but also genes encoding transcription factors such as Activating Transcription Factor 3 (Atf3), Kruppel-like Factor 5 (Klf5) and the transcriptional repressor B-cell Leukemia/Lymphoma 6 (Bcl6). Analysis of conserved sequence elements in promoters of the co-regulated genes identified sites for Signal Transducers and Activators of Transcription (Stats) and for Bcl6 prominently among them. Chromatin immunoprecipitation demonstrated GH-induced reduction in the occupancy of endogenous Bcl6, a transcriptional repressor, which was reciprocal with increasing occupancy of Stat5, on the Socs2 promoter, coincident with GH-stimulated expression of Socs2 mRNA. These studies illustrate the utility of using of multiple computational strategies to predict candidate genes and identify novel transcription factors in the specific biological context.

(* The presented work was mainly done at the University of Michigan before the author moved to the Delaware Biotechnology Institute).

BIOGRAPHY

Poster Session

Turnover in Epithelia is a Mechanism for the Formation of Clonal Multifocal Lesions

[1]**Tom Eck**, [2]Dennis Chao, [2]Georg Luebeck, [3]Carlo Maley

[1]BioMaPS Institute for Quantitative Biology, Rutgers, The State University of New Jersey
Email: teck@cabm.rutgers.edu
[2]Fred Hutchinson Cancer Research Center, Seattle, WA
[3]The Wistar Institute, Philadelphia, PA

ABSTRACT

Multi-focal neoplasms are often observed that are clonally related. Two current hypotheses for this phenomenon are local metastases and expansion of a premalignant clone from which the foci later emerge. Here we introduce a third hypothesis in which clones can fragment due to normal background turnover of cells and tissues. We developed an agent-based model that simulates clonal expansion on a patch of tissue containing crypts arranged in a hexagonal grid. Each crypt possesses a single gene which, when mutated, may confer a selective advantage (reproductive or survival) for that crypt. Crypts die as a result of background turnover or wounding. Mutants with survival advantage are less susceptible to wounding, to a degree controlled by a survival advantage parameter. Crypts divide in a wound healing response to the death of a neighbor after a stochastic waiting time that represents the time required to double the number of cells in the crypt and bifurcate. Mutants with a reproductive advantage benefit from shorter waiting times which is governed by the reproductive advantage parameter. Additional model parameters control the frequency and intensity of wounding. We performed a sweep of the area of parameter space that is most biologically plausible and discovered that the probability of fragmentation is relative to the degrees of reproductive and survival advantage but only for very low fitness or neutral clones. We also found that a clone was most likely to fragment under moderate amounts of wounding and turnover. Similar to disturbance theory in ecology, too little turnover results in the stable, slow growth of the mutant clone. Too much disturbance often results in the extinction of the mutant clone. These results suggest a more parsimonious explanation for clonally related foci of neoplasms than has previously been recognized. Multifocal lesions may emerge from turnover and the expansion of clones with weak fitness advantages over neighboring epithelial cells.

BIOGRAPHY

J. Thomas (Tom) Eck received the B.S. degree in Chemistry with a concentration in Computer Science from Boston College in 1989, the M.S. degree in Computer Science from New Jersey Institute of Technology in 1991, and the M.B. degree in Biotechnology from the University of Pennsylvania in 2005. He is currently a Ph.D. student in the Computational Biology and Molecular Biophysics graduate program at Rutgers, The State University of New Jersey.

Tom Eck's research interests are varied, but focus on the application of interdisciplinary approaches to understand disease mechanism with the goal of developing more effective treatments. While at Penn, Tom studied cancer dynamics and developed an agent-based model of clonal expansion on a tissue. At Rutgers, Tom is primarily focused on the problem of drug resistance in infectious disease and is applying a blend of computational, mathematical, biophysical, and chemical methods to this problem.

Poster Session

Feedback to Descending Projection Neurons Can Override the Mechanisms Underlying Rhythmic Pattern Generation in the Target Network: A Modeling Study

Nickolas Kintos

Visiting Assistant Professor, Department of Mathematics
Fordham University
New York, NY 10023
Tel: 212-636-6333, Fax: 212-636-7217
Email: kintos@fordham.edu

ABSTRACT

Rhythmic motor networks are generally studied assuming feed-forward architecture. Yet, feedback to descending inputs is pervasive in motor pattern generation, but its role is not well understood. We developed a mathematical model to understand how rhythmic feedback to a descending projection neuron shapes the output of the gastric mill (chewing) motor circuit in the stomatogastric ganglion (STG) of the crab *C. borealis*. Stimulating the projection neuron MCN1 elicits a gastric mill rhythm (GMR) in vitro, where the protractor LG neuron bursts in antiphase with the retractor neuron INT1 (Coleman et al, 1995 Nature). The half-center oscillation of the reciprocally inhibitory LG-INT1 pair underlies the MCN1-elicited GMR. Co-activation of MCN1 and the projection neuron CPN2 elicits a distinct GMR, where the CPN2 soma is inhibited by a feedback synapse from INT1. Previous experiments indicated that the MCN1/CPN2-GMR (but not the MCN1-GMR) persists without the inhibitory synapse from INT1 to LG.

We use a reduced mathematical model in parallel with a biophysically-detailed model to elucidate the mechanisms that enable the MCN1/CPN2-GMR (but not the MCN1-GMR) to persist without INT-LG reciprocal inhibition. We show that INT1 feedback to CPN2 causes the MCN/CPN2-GMR to persist without the INT1 to LG inhibitory synapse. In contrast, the MCN1-GMR is disrupted without that synapse. We conclude that network feedback to projection neurons can move the locus of pattern generation from a half-center oscillator to an excitation-feedback circuit, which in turn can alter inter-circuit interactions. (Joint work with Professors Farzan Nadim - NJIT; Michael Nusbaum - U Penn School of Medicine).

BIOGRAPHY

Nickolas Kintos received his B.A. in Mathematics and B.A. in Chemistry from Rutgers University (Newark, NJ) in 2000. He received his M.S. in Applied Mathematics from the New Jersey Institute of Technology (NJIT – Newark, NJ) in 2005. He received his Ph.D. in Mathematical Sciences from NJIT in 2007. Currently he is a Peter M. Curran Visiting Assistant Professor in the Department of Mathematics at Fordham University (New York, NY).

Nickolas Kintos' research interests include singularly perturbed dynamical systems and computational neuroscience. In particular, he builds mathematical models, based upon the experimental data of collaborators, to investigate the dynamics that underlie the activity of neural networks. He is currently a member of the Society for Neuroscience (SfN) and the Society for Industrial and Applied Mathematics (SIAM). He has recently co-authored a research paper (Journal of Computational Neuroscience) and several abstracts for conference proceedings (Society for Neuroscience Meeting, SIAM Conference on the Life Sciences, and East Coast Nerve Net – Marine Biological Laboratories, Woods Hole, MA).

Poster Session

Detecting Mutation Rate Change in Barrett's Esophagus after Treatment with NSAIDs

Rumen Kostadinov[1], Mary Kuhner[2], Carlo Maley[1]

[1]Genomics and Computational Biology
University of Pennsylvania
Philadelphia, PA 19104
Email: rumen@mail.med.upenn.edu

[2]Department of Genome Sciences
University of Washington
Seattle, WA 98195-7360

ABSTRACT

Non-steroidal anti-inflammatory drugs (NSAIDs) have been shown to reduce the risk of esophageal adenocarcinoma (EA) in Barrett's Esophagus (BE) patients (Vaughan et al., Lancet Oncol. 2005;6(12):945-52). However, the mechanism through which NSAIDs delay progression to EA is not well understood.

We hypothesize that NSAID use slows clonal evolution by reducing mutation rate. To test this hypothesis in-vivo, Illumina SNP arrays can be used to detect loss of heterozygosity (LOH) events in frozen longitudinal biopsies from BE patients from the Seattle Barrett's Esophagus cohort. Using coalescent theory, the observed LOH events, known sampling times and known onset of NSAIDs treatment we can estimate the rate of LOH per year, the stem cell population size, and changes in these parameters after treatment.

We performed in-silico analysis of the power to detect mutation rate changes, which is critical for selecting the number of time points and biopsies per time point to assay per patient. We modified SerialSimCoal (Excoffier et.al. J. Hered., 91, 506-509) to simulate a scenario in which 4 biopsies were taken from patients every 2 years for 10 years, with NSAID use beginning in year 5, and asked whether we could reliably detect mutation rate changes of various intensities (1.1, 2, 5, 10, and 100-fold). We varied stem cell population sizes and base mutation rates over 3 orders of magnitude on DNA sequences of 8,000 loci. We used BEAST (Drummond et.al. Genetics 161, 1307-1320) to estimate population sizes and mutation rates before and after year 5.

For 91% of parameter combinations, we had >80% power to detect a 1.1- to 100- fold decrease in mutation rate (paired Wilcoxon test, $p<0.05$). These simulations allow for the design of longitudinal studies to be carried out in BE cohorts to determine the in-vivo effects of NSAIDs.

BIOGRAPHY

Poster Session

A framework for discovering associations from the annotated biological web

Woei-jyh Lee

Center for Bioinformatics and Computational Biology, and
Institute for Advanced Computer Studies, and
Department of Computer Science
University of Maryland
College Park, Maryland 20742

Email: adamlee@umiacs.umd.edu

ABSTRACT

During the last decade, biomedical researchers gained access to the entire human genome, reliable high-throughput biotechnologies, and affordable computational resources and network access. In combination, these new tools created a new model for biomedical research that no longer uses computational tools merely to monitor research, but instead exploits these tools to acquire knowledge and make discoveries. Consider a simplified web of three publicly accessible resources Entrez Gene, OMIM and PubMed. Data entries in each resource are annotated with terms from multiple controlled vocabularies (CVs). The hyperlinks between data entries in any two resources form a relationship between the two resources and is represented by a (virtual) link. Thus, an entry in Entrez Gene, annotated with GO terms, can have hyperlinks to multiple entries in PubMed that are annotated with MeSH terms. Similarly, OMIM entries, annotated with terms from SNOMED may have hyperlinks to entries in Entrez Gene and PubMed. This forms a rich web of annotated data entries. Our objective in this research is to develop tools to discover meaningful patterns across resources and ontologies. As a first stage in teasing out patterns, we execute a protocol to follow hyperlinks, extract annotations, and generate LSLink datasets. We then mine the term-links of the LSLink datasets to find potentially meaningful associations. Biologically meaningful associations of pairs of CV terms may yield actionable nuggets of previously unknown knowledge. Moreover, the bridge of associations across CV terms will reflect the practice of how scientists annotate data across hyperlinked repositories.

This is a joint work with Louiqa Raschid, Hassan Sayyadi and Padmini Srinivasan.

BIOGRAPHY

Woei-jyh (Adam) Lee received his B.S. degree from the Department of Computer Science and Information Engineering at the National Taiwan University in 1993, and his M.S. degree from the Department of Computer Science at the New York University in 1998.

He worked on distributed objects and fault tolerance at AT&T Labs - Research in 1997. He focused on network software and management at Bell Laboratories Research, Lucent Technologies, from 1998 till 2000. He visited Integrated Media Systems Center at the University of Southern California specializing in continuous media streaming and multimedia networking from 2002 to 2003. He also contributed in protein domain parsing and boundary prediction at the National Cancer Institute, National Institutes of Health. He is currently a PhD candidate in the Department of Computer Science at the University of Maryland at College Park. He is also affiliated with the Center for Bioinformatics and Computational Biology and the Institute for Advanced Computer Studies at the University of Maryland. His research interests include bioinformatics, computational biology, systems biology, genomics and genetics, data management, information integration and data mining.

Mr. Lee is a member of the ACM, the IEEE, the ISCB and the ISENG.

Poster Session

A Markov Random Field-based Edit Distance (MRFED) Algorithm for Gene Synonym Matching

Min Song

Assistant Professor
Information Systems department, GITC 4201
New Jersey Institute of Technology
323 Martin Luther King, Jr. Blvd., Newark, NJ 07102
Email: min.song@njit.edu

ABSTRACT

The edit distance has played important roles in a wide array of applications due to its representational efficacy and computational efficiency. There is, however, a fundamental problem with the original definition of edit distance that has remained elusive: its context-free nature. In determining the possible actions, i.e., insertion, deletion, and substitution, no systematic consideration was given to the local behaviors of the string/pattern in question that indeed encompass great amount of useful information regarding its content. In this poster, we propose a new edit distance algorithm, called MRFED, by applying the Markov Random Field theory to Needleman-Wunch distance for gene synonym matching. The formula of MRFED is as follows:

$$MRFED_{i,j}\begin{cases}\dfrac{NW_{ij}}{L(P_{ij})}, & if \quad h(A[i-k+1,i]) \neq h(B[j-k+1,j]) \\ \min\left(\dfrac{NW_{ij}}{L(P_{ij})}, \dfrac{MRFED_{i-k,j-k}-0.5\times(k-1)}{L(P_{ij})}\right) & otherwise \end{cases} \quad ,k=1..K$$

Where 1) NW_{ij} – Needleman-Wunch Distance for two substrings A[1...i] and B[1...j]; 2) L(Pij) - length of the edit path; 3) h() - histogram function; 4) A[i-k+1,i] - substring of A starting from (i-k+1)th symbol and ending with i-th symbol; 5) B[j-k+1,j]- substring of B starting from (j-k+1)th symbol and ending with j-th symbol; 6) MRFEDi-k, j-k - the value of MRFED for two substrings A[1...(i-k)] and B [1..(j-k)]; and 7) small k - a parameter changing from 1 to the capital K which is the order of neighborhood system.

It is an important step to find the unique identifier of gene (a.k.a geneId finding) in the curation process. GeneId finding is the task of finding the database identifier of every gene discussed in an article, and it was studied experimentally in the BioCreatIvE challenge (Hirschman et al., 2005), which developed test-bed problems for each of three model organisms (yeast, mice, and fruitflies). We focused on the mouse dataset, which was the hardest for the BioCreatIvE participants. This dataset consists of several parts. The gene synonym list consists of 183,142 synonyms for 52,594 genes. Here we consider gene synonym matching, the matching task of gene synonyms.

Algorithm	**Measure**	
	Average Precision	**MaxF1**
SmithWaterman	0.44	0.55
SoftTFIDF	0.62	0.67
Levenstein	0.35	0.31
Jaccard	0.54	0.59
MRFED	0.59	0.61
SoftMRFED	**0.67**	**0.72**

Table 1: Performance Results for Gene Synonym Matching with Mice Datasets

We compare our MRFED algorithm with other string similarity algorithms such as softTFIDF, Levenstein, and Jaccard. By integrating our MRFED algorithm with a token-based similarity, called SoftMRFED, we further improve the performance of the technique. As shown in Table 1, our SoftMRFED outperforms the other algorithms while our original MRFED performs equivalent to others.

BIOGRAPHY

Min Song was born in Suwon, Korea in July 23rd, 1969. is an assistant professor of Department of Information Systems at NJIT. He received his M.S. in School of Information Science from Indiana University in 1996 and received Ph.D. degree in Information Systems from Drexel University in 2005.

He worked at Thomson Scientific for eight years as a senior software engineer from 1999 to 2005. He is currently an assistant professor in Information Systems at New Jersey Institute of Technology. He has a background in Text Mining, Bioinfomatics, Information Retrieval and Information Visualization.

Dr. Song received the Drexel Dissertation Award in 2005. In 2006, Min's work received an honorable mention award in the 2006 Greater Philadelphia Bioinformatics Symposium. In addition, The paper entitled "Extracting and Mining Protein-protein interaction Network from Biomedical Literature" has received the best paper award from 2004 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, which was held in San Diego, USA, Oct. 7-8, 2004. In addition, another paper entitled "Ontology-based Scalable and Portable Information Extraction System to Extract Biological Knowledge from Huge Collection of Biomedical Web Documents" was nominated as the best paper at 2004 IEEE/ACM Web Intelligence Conference, which was held in Beijing, China, Sept, 20-24, 2004.

Poster Session

Towards an automatic classification of protein structural domains based on structural similarity

Vichetra Sam[1], **Chin-Hsien Tai**[2], Jean Garnier1[3], Jean-Francois Gibrat[3], Byungkook Lee[2], Peter J. Munson[1]

[1] Mathematical and Statistical Computing Laboratory, CIT, NIH, Bethesda, MD, USA,
[2] Laboratory of Molecular Biology, NCI, NIH, Bethesda, MD, USA and
[3] Mathematique Informatique et Genome, INRA, Jouy-en-Josas, France
Email: taic@mail.nih.gov

ABSTRACT

Understanding the protein structure universe and establishing evolutionary relationships among them are aided by classification. Well-known classification databases, such as SCOP and CATH, are manually curated at some stage and are different from automatic classifications such as FSSP. The difference could arise from the pair-wise similarity scores used in automatic classification. But another possible reason is the clustering procedure used. Here, we explore the degree to which these two factors might affect the final classification.

We used DALI, SHEBA and VAST pairwise structural similarity measures on the ASTRAL40 SCOP C class domains, to investigate a variety of clustering procedures. All procedures first construct a dendrogram, which hierarchically joins domains into progressively less similar groups. The dendrogram is then cut in a variety of ways to produce a partition, which is compared to the SCOP Fold classification.

We found that no procedure exceeded an average of 61% true positives at the 1% false positive rate (FPR). Even procedures that explicitly optimize similarity to SCOP gave an average of only 72% true positives at the 1% FPR. We also developed a method to detect differences that are inherent in the dendrogram and are, therefore, irreducible by any dendrogram-cutting strategies. We found that these irreducible differences are substantial.

These observations, as well as visual examination of individual cases, indicate that the major difference between automatic and manual protein classification arises from the pairwise scores themselves, rather than from the clustering procedure. This is a strong call for the need for an improved structure comparison programs/protocols.

BIOGRAPHY

Chin-Hsien (Emily) Tai received her BS degree from Department of Zoology at National Taiwan University in 1994, her MS degree from Graduate Institute of Immunology at National Taiwan University in 1996 and another MS degree in Computer and Information Science from New Jersey Institute of Technology in 1999. She worked in Information Technology Division in Goldman, Sachs & Co. in New York and London from 1999 to 2002. She joined Molecular Modeling and Bioinformatics Section, Laboratory of Molecular Biology at National Cancer Institute, NIH as a staff scientist since 2002. Her research interests include bioinformatics and protein structure analysis.

務實服務

科技交流

- 促進台灣與美國間學術科技交流合作
- 倡導科技產業投資及技術轉移
- 加強台灣與美國產、官、研、學界互訪
- 延攬海外科技人才
- 協助海外科技社團舉辦科技活動
- 協助國內蒐集科技資訊
- 服務區域包括大華府地區、美東地區、美東北地區、美東南地區及美中地區共二十四州

歡迎利用各項服務

Taipei Economic & Cultural Representative Office in the United States
Science and Technology Division
4201 Wisconsin Avenue NW, MB-09
Washington, DC 20016-2137
Tel: 202-895-1930, Fax: 202-895-1939, 網址：http://dc.nsc.gov.tw
E-Mail: dc@tecrosd.org

TECRO Science and Technology Division

駐美國台北經濟文化代表處科技組