The Sixth IEEE International Conference on Systems Biology (IEEE ISB 2012)

Local Organizer

APORC
Asia-Pacific
Operations Research Center

Organizers

Sponsors

August 18-20, 2012
Xi'an, Shanxi, China

ISB2012 Sessions Locations

Date		Conference room Peony	Conference room Pingpong
August 17 Friday	**15:00-23:00**	Registration	
	18:00-19:00	Dinner (Rongyuan restaurant)	
August 18 Saturday	**Morning 08:00-12:20**	ISB P1 ISB P2	
	Afternoon 14:00-18:30	ISB H1 ISB H2	ISB B1 ISB B2
	Evening 18:30-19:30	Welcome reception(Conference room Lily)	
	Evening 19:30-21:00	Board member meeting for Computational Systems Biology Society	
August 19 Sunday	**Morning 08:00-12:30**	ISB H3 ISB H4	ISB B3 ISB B4
	Afternoon 14:00-18:30	ISB H5	ISB B5
		ISB A1	ISB B6
	Evening 18:30-21:00	Tang dinner, departure at 18:30 from lobby	
Augst 20 Monday	**Morning 08:00-12:30**	ISB A2 ISB A3	ISB B7 ISB B8
	Afternoon 13:30-18:30	Half day excursion in The Terra Cotta Warrior & House Museum.Departure at 13:20 from lobby	
	Evening 18:30-21:00	Banquet	

ISB2012 Schedule

August 17 Friday	15:00-23:30	Registration (hotel lobby at Xi'an Jianguo Hotel)	
	18:00-19:00	Dinner (Rongyuan Chinese Restaurant)	
August 18 Saturday	08:00-08:30	Opening Session (Conference room Peony, floor-1)	
	08:30-10:10	ISB Plenary Session P1	
	10:10-10:40	Coffee break	
	10:40-12:20	ISB Plenary Session P2	
	12:30-14:00	Lunch (Rongyuan restaurant)	
	14:00-16:05	ISB Session H1 (Conference room Peony, floor-1)	ISB Session B1(Pingpong Room)
		Highlight: Computational Genomics	Regular: Network Biology
	16:05-16:25	Coffee break	
	16:25-18:30	ISB Session H2 (Conference room Peony)	ISB Session B2 (Pingpong room)
		Highlight: Proteomics	Regular: Dynamics Systems Biology
	18:30-19:30	Reception (Conference room Lily)	
	19:30-21:30	Board member meeting for Computational Systems Biology Society (location to be announced)	
August 19 Sunday	08:00-10:05	ISB Session H3 (Conference room Peony)	ISB Session B3(Pingpong room)
		Highlight: Bioinformatics	Regular: Network Biology
	10:05-10:25	Coffee break	
	10:25-12:30	ISB Session H4 (Conference room Peony)	ISB Session B4(Pingpong room)
		Highlight: Bioinformatics	Regular: Next Generation Data Analysis
	12:30-14:00	Lunch (Buffet at Xi'an Jianguo Hotel)	
	14:00-16:05	ISB Session H5 (Conference room Peony)	ISB Session B5(Pingpong room)
		Highlight/Regular: Bioinformatics	Regular: Bioinformatics
	16:05-16:25	Coffee break	
	16:25-18:30	ISB Session A1 (Conference room Peony)	ISB Session B6(Pingpong room)
		Regular:Complex disease	Regular:Dynamics Systems Biology
	18:30-20:00	Tang dinner, departure at 18:30 from lobby	

ISB2012 Schedule

August 20 Monday	08:00-10:05	ISB Session A2 (Conference room Peony)	ISB Session B7(Pingpong room)
		Regular: Integrative Bioinformatics	Regular:Dynamics Systems Biology
	10:05-10:25	Coffee break	
	10:25-12:30	ISB Session A3 (Conference room Peony)	ISB Session B8(Pingpong room)
		Regular: Bioinformatics	Regular:Bioinformatics
	12:30-13:30	Lunch (Rongyuan restaurant)	
	13:30-18:00	Half day excursion in The Terra Cotta Warrior & House Museum.Departure at 13:20 from lobby	
	18:30-22:00	Banquet, Xi'an Grant Restaurant	

Parallel Session: August 18, Afternoon (14:00pm--18:30pm)

ROOM→ Time↓	Conference room Peony, floor-1)	Pingpong Room
Parallel Sessions	**Highlights Track**	**Proceedings Track**
14:00-14:25	*Predicting nuclear export regulatory signals from amino acid sequence* **Paul Horton**	*Hierarchical Modular Structure in Gene Co-expression Networks* **Shuqin Zhang**
14:25-14:50	*A fast and accurate SNP detection algorithm for next generation sequencing data* **Junwen John Wang**	*A Gaussian Graphical Model for Identifying Significantly Responsive Regulatory Networks from Time Series Gene Expression Data* **Wanwei Zhang**
14:50-15:15	*Wavelet-based identification of DNA focal genomic aberrations from single nucleotide polymorphism arrays* **Hyunju Lee**	*A Seed-based Approach to Identify Risk Disease Sub-networks in Human Lung Cancer* **Yi-Bin Wang**
15:15-15:40	*Exploiting a Reduced Set of Weighted Average Features to Improve Prediction of DNA-binding Residues from 3D Structures* **Juan Liu**	*Pigmented Network Structure Detection Using Semi-Smart Adaptive Filters* **Leszek A. Nowak**
15:40-16:05	*PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites* **Jiangning Song**	*ppiPre - an R package for predicting protein-protein interactions* **Yue Deng**
16:05-16:25 Coffee break		
16:25-16:50	*Systematic analysis of the Plk-mediated phosphoregulation in eukaryotes* **Yu Xue**	*The influence of the basin structure of Boolean networks on their long range correlated dynamics* **Wenbin Liu**
16:50-17:15	*Systematic analysis and prediction of longevity genes in Caenorhabditis elegans* **Yan-Hui Li**	*Functional tunability of biological circuits from tinkers* **Changhong Shi**
17:15-17:40	*Unrestrictive Identification of Protein Post-translational Modifications from Tandem Mass Spectra* **Yan Fu**	*Escape from infinite adaptive peak* **Song Xu**
17:40-18:05	*Comparative Analysis of Different Label-Free Mass Spectrometry Based Protein Abundance Estimates and Their Correlation with RNA-Seq Gene Expression Data* **Ning Kang**	*A Stable Simplification of a Fas-signaling Pathway Model for Apoptosis* **Ya-Jing Huang**

18:05-18:30	*Bridging modern and traditional Chinese medicines by knowledge-mining* **Xiaoqin Yang**	*Switch-Like Regulation of Signal Transduction by Small RNA-mediated Quorum Sensing* **Xi Liu**

Parallel Session: August 19, Morning (08:00am--12:30pm)

ROOM→ Time↓	Conference room Peony, floor-1)	Pingpong Room
Parallel Sessions	**Highlights Track**	**Proceedings Track**
08:00-08:25	*Network-based approach for gene set analysis* **Weidong Tian**	*Construction and Analysis of Genome-wide SNP Networks* **Michael K. Ng**
08:25-08:50	*Evolution and genetic variation of microRNA mediated gene regulation in humans* **Zhaolei Zhang**	*Effective Clustering of MicroRNA Sequences by N-grams and Feature Weighting* **Shuigeng Zhou**
08:50-09:15	*Discovery of multi-dimensional modules by integrative analysis of cancer genomic data* **Shihua Zhang**	*Network Clustering along Diabetes Progression in Three Tissues of Goto-Kakizaki Rats* **Katsuhisa Horimoto**
09:15-09:40	*MetaBinG: Using GPUs to Accelerate Metagenomic Sequence Classification* **Chaochun Wei**	*Network Kernel SVM for Microarray Classification and Gene Sets Selection* **Bing Yang**
09:40-10:05	*Meta-Storms: Efficient Search for Similar Microbial Communities Based on a Novel Indexing Scheme and Similarity Score for Metagenomic Data* **Xiaoquan Su**	*cGRNexp: a Web Platform for Building Combinatorial Gene Regulation Networks based on user-uploaded gene expression datasets* **Hui Yu**
10:05-10:25 Coffee break		
10:25-10:50	*Modeling and comparing the organization of circular genomes* **Grace S. Shieh**	*RNA-seq Coverage Effects on Biological Pathways and GO Tag Clouds* **Chien-Ming Chen**
10:50-11:15	*Efficient Exponential Time Algorithms for Edit Distance between Unordered Trees* **Takeyuki Tamura**	*Statistical distribution of transcription (gene expression) in eukaryotic cells: insight from RNA-seq data.* **Nelson Tang**
11:15-11:40	*GWASdb: a database for human genetic variants identified by genome-wide association studies* **Mulin Jun Li**	*A machine learning framework of functional biomarker discovery for different microbial communities based on metagenomic data* **Wei Fang**
11:40-12:05	*Efficient methods for identifying mutated driver pathways in cancer* **Junfei Zhao**	*BAsplice: Bi-direction Alignment for detecting splice junctions* **Jingde Bu**

12:05-12:30	*Using Graphical Adaptive Lasso approach to construct TF and miRNA's combinatorial regulatory network in breast cancer* **Naifang Su**	*Identifying Mutated Core Modules in Glioblastoma by Integrative Network Analysis* **Junhua Zhang**

Parallel Session: August 19, Afternoon (14:00pm--18:30pm)

ROOM→ Time↓	Conference room Peony, floor-1)	Pingpong Room
Parallel Sessions	**Proceedings Track**	**Proceedings Track**
14:00-14:25	*Understanding Virus Infection and Therapy based on Mathematical Modeling and Experiments* **Xiao Chen**	*Comparative analysis of protein-coding genes and long non-coding RNAs of prostate cancer between Caucasian and Chinese populations* **Yan Zhang**
14:25-14:50	*CNetA: Network alignment by combining biological and topological features* **Qiang Huang**	*A Sequence-Segmented Method Applied to the Similarity Analysis of Proteins* **Ping-an He**
14:50-15:15	*FALCON ---- predicting protein structure prediction using fragment-HMM with an optimized energy function* **Dongbo Bu**	*New encoding schemes for prediction of protein Phosphorylation sites* **Zimo Yin**
15:15-15:40	*A 3-Dimentional Multiscale Model to Simulate Tumor Progression in Response to Interactions between Cancer Stem Cells and Tumor Microenvironmental Factors* **Ming Zhan**	*Analysis of Morphological Evolution in a Long-term Experiment with {\it Escherichia coli* **Fangshu Cui**
15:40-16:05	*Detecting early-warnings signal for complex diseases by quantifying conditional network entropy* **Rui Liu,**	*Predicting Protein-RNA Residue-base Contacts Using Two-dimensional Conditional Random Field* **Morihiro Hayashida**
16:05-16:25 Coffee break		
16:25-16:50	*Using NMFAS to Identify Key Biological Pathways Associated With Human Diseases* **Hao Guo**	*Multi-objective Optimization of Biological Systems Represented by S-system Models* **Gongxian Xu**
16:50-17:15	*Clinical Data Analysis Reveals Three Subytpes of Gastric Cancer*	*System identification of the fermentation system of*

	Yong Wang	*Thermoanaerobacter sp. X514* **Jing Yang**
17:15-17:40	*in silico identification of novel cancer-related genes by comparative genomics of naked mole rat and rat* **Zhiyuan Yang**	*Coupled Positive Feedback Loops Regulate the Biological Behavior* **Fei Shi**
17:40-18:05	*Metabolite Biomarker Discovery For Metabolic Diseases By Flux Analysis* **Hao Jiang**	*Alternating Weighted Least Squares Parameter Estimation for Biological S-Systems* **Fang-Xiang Wu**
18:05-18:30	*Identification of Oncogenic Genes for Colon Adenocarcinoma from Genomics Data* **Changhe Fu**	*Dynamics of Coexistence of Asexual and Sexual Reproduction in Adaptive Landscape* **Shuyun Jiao**

Parallel Session: August 20, Morning (08:00am--12:30pm)

ROOM→ Time↓	Conference room Peony, floor-1)	Pingpong Room
Parallel Sessions	**Proceedings Track**	**Proceedings Track**
08:00-08:25	*Improving Prediction of Drug Therapy Outcome via Integration of Time Series Gene Expression and Protein Protein Interaction Network* **Liwei Qian**	*Comparing two models based on the transcriptional regulation by KaiC of cyanobacteria rhythm* **Ying Li**
08:25-08:50	*An Integrative Framework for Identifying Consistent MicroRNA Expression Signatures Associated with Clear Cell Renal Cell Carcinoma* **Bairong Shen**	*New Global Stability Conditions for Genetic Regulatory Networks with Time-Varying Delays* **Li-Ping Tian**
08:50-09:15	*Identifying novel glioma associated pathways based on integrated 'omics' data* **Guang Hu**	*Module network rewiring in response to therapy* **Tao Zeng**
09:15-09:40	*A Novel Pipeline for Motif Discovery, Pruning and Validation in Promoter Sequences of Human Tissue Specific Genes* **Xiu-Jun Gong**	*Dynamic miRNA-TF-mRNA circuits in mouse lung development* **Juan Liu**
09:40-10:05	*Predicting protein complexes via the integration of multiple biological information*	*Analysis of A HBV Infection Models with ALT* **Yongmei Su**

	Jianxin Wang	
10:05-10:25 Coffee break		
10:25-10:50	*Human Encoded miRNAs that Regulate the Inflenenza Virus Genome* **Hao Zhang**	*A New Method to Identify Repositioned Drugs for Prostate Cancer* **Zikai Wu**
10:50-11:15	*A Novel Information Contents Based Similarity Metric for Comparing TFBS Motifs* **Shaoqiang Zhang**	*Application of Granger Causality to Gene Regulatory Network Discovery* **Gary Hak Fui Tam**
11:15-11:40	*A fixed-point blind source extraction algorithm and its application to ECG data analysis* **Zikai Wu**	*A Novel Feature Selection Method Based on CFS in Cancer Recognition* **Yong Deng**
11:40-12:05	*Anti-clustering of circadian gene expression in mouse liver genome* **Yuan-yuan Li**	*Sparse Kernel Logistic Regression for β-turns Prediction* **Jianxin Wang**
12:05-12:30	*Predicting biological dynamics based on short high throughput time-course data* **Huanfei Ma**	

IEEE ISB 2012 Program

August 18-20, Xi'an, Shanxi, China

August 17 (Friday) Registration

15:00-23:30 Registration, Participants arrival in Xi'an, check in Xi'an Jianguo Hotel, and Registration package pick up (Hotel Lobby at Xi'an Jianguo Hotel).

18:00-19:00 Dinner (Xi'an Jianguo Hotel)

19:00-21:30 Board member Meeting for Computational Systems Biology Society of ORSC (Conference Room Pingpong in Xi'an Jianguo Hotel)

August 18 (Saturday) Technical sessions

07:30-11:30 Registration for late arrivals *(*Hotel Lobby at Xi'an Jianguo Hotel*)*

08:00-08:30 Opening Session for ISB2011 (Conference Room Peony at Xi'an Jianguo Hotel)

Chair: Luonan Chen

8:30-10:10 ISB Plenary Session P1 (Conference Room Peony at Xi'an Jianguo Hotel)

Chair: Luonan Chen

8:30-9:20 *Swine, avian and human influenza and the origins of influenza A (H1N1) pandemic strain*

Raul Rabadan

Center for Computational Biology and Bioinformatics, Columbia University College of Physicians and Surgeons, USA

Abstract: Evolution is a dynamical process that shape the genomes of viruses and their hosts. Large amounts of genomic data are currently available allowing the identification of patterns in mutations and reassortments, selection and pathogenesis. In particular, we will explore these evolutionary patterns in the context of the 2009 influenza A (H1N1) pandemic strain.

9:20-10:10 *Synthetic biology for the comprehension of biomolecular networks*

Masahiro Okamoto

Department of Bioinformatics, Graduate School of Systems Life Sciences, Kyushu University

Abstract: In order to make the paradigm shift from the concept of "watched and analyzed biology" to that of "synthetic and analyzed or utilized biology", the innovative research named Synthetic Biology was started from 2000 in US, such as designing synthetic genetic circuit by combining known interrelated biomaterials, realizing a certain bio-functional behaviors such as switch, oscillation in vivo, designing synthetic metabolic pathways by incorporating enzyme coded genes from other origins into the cells. However, these attempts have been done on a small scale and with a trial-and-error method. The objectives of this research project is to establish the coordination between the fundamental technologies for synthetic biology in order to comprehend biomolecular networks by integrating the following three missions: 1) design synthetic genetic circuit or metabolic pathway with using the methods of computational science, 2) construct the circuit in vitro with using the method of engineering, 3) construct the circuit in vivo or in the cell with using the methods of molecular biology. In order to construct and control a large scale of dynamic and complex synthetic genetic circuit or metabolic pathways, the following fundamental technologies for synthetic biology are essential: Biochemical Engineering, Embryological Engineering, Molecular Biology, Evolutional

Molecular Engineering, Micro Fluid Engineering, Biomolecular Chemistry, Simulation Engineering, and Knowledge-based Engineering. In the first stage (2-3 years), our mission is to construct dynamic and multi-elements synthetic genetic circuit, followed by the construction of differentiation-induced system against stem cell and by the realization of cell factory, in which cells can produce the target metabolites by themselves according to the cell environment in the last 2-years. The research project is composed of the following four sub sections: (A01) fundamental technologies of molecular biology, (B01) fundamental technologies of engineering, (C01) fundamental technologies of computational science, and (X01) integrated section of A01, B01 and C01.

10:10-10:40 Coffee break

10:40-12:20 ISB Plenary Session P2 (Conference Room Peony at Xi'an Jianguo Hotel)

Chair: Lin Gao

10:40-11:30 *From Structure Based to Systems Based Drug Design*

Luhua Lai

College of Chemistry and Molecular Engineering, & Center for Quantitative Biology, Peking University, China

Abstract: Structural based drug design has been widely used in drug discovery for leading compounds identification and optimization. Many successful applications have been reported. Various docking methods and de novo structural based drug design programs have been developed, which were all developed based on the single target binding assumption. However, drugs encounter complicated situations and a huge number of biological molecules in the human body and may cause unexpected deleterious or beneficial effects. In order to understand mechanism of drug action, disease related molecular networks need to be studied. My group has been working on the human inflammation related arachidonic acid (AA) metabolic network in order to undertand its regulation mechanism and to discover better intervention strategy. We have developed a multiple target optimum intervention method (MTOI) that can be used to identify key targets for drug design and to predict optimum solutions to modulate disease network with minimum side effects. MTOI has been successfully used in simulating the inflammation related arachidonic acid metabolic network and in predicting novel combinatorial intervention strategies. As systems-based drug intervention often requires simultaneous control of multiple targets, we also explored possible ways for multi-target based drug design. Several types of multiple target inhibitors were discovered in the AA network and their influences on the AA metabolic network were studied. Methods and strategies developed in the AA metabolic network study can be generally applicable in studying other disease related networks.

11:30-12:20 *Genome-wide protein structure prediction and structure-based function annotation*

Yang Zhang

Department of Computational Medicine and Bioinformatics, Department of Biological Chemistry, University of Michigan, USA

Abstract: The biological function of protein molecules is determined by the shape of their three-dimensional structures. Is it possible to predict protein structure and function from the amino acid sequence? We developed a new algorithm, I-TASSER, which assembles atomic structure of proteins using fragments excised from unrelated experiment structures. Functional insights (e.g. ligand-binding affinity, enzyme classification and gene ontology) are then deduced by matching the predicted structure models with the known proteins in protein function libraries. The I-TASSER algorithm was ranked as the best for automated protein structure prediction in the communitywide CASP experiments of 2006, 2008 and 2010; it was also ranked at the top for protein function annotation in CASP9 in 2010.

In this talk, I first review recent progress in computer-based protein structure prediction including the new developments in ab initio folding and atomic structure refinements since the CASP9 experiment, and show that the protein structure prediction problem can in principle be solved using the current PDB library. Next, we discuss the application of the developed methods to the structural and functional modeling of a number of genomes, including all G-protein coupled receptors (GPCRs) in the human genome, yielding models 90% of which are shown to have correct topology, and Marek's disease virus, the first success of the computational modeling of a complete viral genome. Finally, we demonstrate how the predicted I-TASSER structure models can be used to annotate the biological function of the proteins and screen drug candidates by matching their global topology and functional sites against the existing structure/function/binding databases.

12:30-14:00 Lunch break (Rongyuan Restaurant at Xi'an Jianguo Hotel)

14:00-16:05 ISB Highlight Session H1 (Conference Room Peony at Xi'an Jianguo Hotel)

Topic: Computational Genomics
Chair: Shihua Zhang

14:00-14:25 *Predicting nuclear export regulatory signals from amino acid sequence*

Paul Horton

University of Tokyo & AIST, Computational Biology Research Center, Japan

Abstract: The classical (leucine-rich) nuclear export signal (NES), is a protein localization signal often involved in important processes such as signal transduction and cell cycle regulation. Although 15 years has passed since its discovery, limited structural information and high sequence diversity have hampered understanding of the NES. Several consensus sequences have been proposed to describe it, but they suffer from poor predictive power. On the other hand, the NetNES server provided the only computational method available. Although these two methods have been widely used to attempt to find the correct NES position within potential NES-containing proteins, their performance had not yet been evaluated on the basic task of identifying NES-containing proteins. We proposed a new predictor, NESsential, which uses sequence derived meta-features, such as predicted disorder and solvent accessibility, in addition to primary sequence. We demonstrated that it can identify promising NES-containing candidate proteins (albeit at low coverage), but other methods cannot. We also quantitatively demonstrated that predicted disorder is a useful feature for prediction and investigate the different features of (predicted) ordered versus disordered NES's. Finally, we listed 70 recently discovered NES-containing proteins, doubling the number available to the community.

As a follow-up to this published research, we gathered another 76 validated NES training examples and prepared a web server to conveniently browse the dataset and make new predictions http://validness.ym.edu.tw/. The source code for NESsential is open source and can be downloaded from http://seq.cbrc.jp/NESsential/.

The talk is based on our papers:

Szu-Chin Fu, Kenichiro Imai and Paul Horton. Prediction of leucine-rich nuclear export signal containing proteins with NESsential. Nucleic Acids Research, doi:10.1093/nar/gkr493, 2011. [Faculty of 1000]

Szu-Chin Fu, Hsuan-Cheng Huang, Paul Horton, Hsueh-Fen Juan. ValidNESs: an integrated, up-to-date database and prediction tool for proteins containing leucine-rich nuclear export signal Under preparation for submission to NAR 2012 database issue

14:25-14:50 *A fast and accurate SNP detection algorithm for next-generation-sequencing data*

Junwen John Wang

Department of Biochemistry & Centre for Genomic Sciences, The University of Hong Kong, China

Abstract:Tools have been developed to call Single Nucleotide Polymorphism (SNP) based on Next-generation-sequencing (NGS) data. However, most of them require high sequencing depth, which is too expensive to obtain. Here, we propose a fast and accurate SNP-detection program, FaSD, which could finish SNP calling within four hours for ten-fold human genome NGS data (total 30 gigabases (GB)) in a standard desktop computer with high accuracy.

14:50-15:15 *Wavelet-based identification of DNA focal genomic aberrations from single nucleotide polymorphism arrays*

Hyunju Lee

Dept of Information and Communications, Gwangju Institute of Science and Technology, Gwangju, South Korea

Abstract: Copy number aberrations (CNAs) are an important molecular signature in cancer initiation, development, and progression. However, these aberrations span a wide range of chromosomes, making it hard to distinguish cancer related genes from other genes that are not closely related to cancer but are located in broadly aberrant regions. With the current availability of high-resolution data sets such as single nucleotide polymorphism (SNP) microarrays, it has become an important issue to develop a computational method to detect driving genes related to cancer development located in the focal regions of CNAs. In this study, we introduce a novel method referred to as the wavelet-based identification of focal genomic aberrations (WIFA). The use of the wavelet analysis, because it is a multi-resolution approach, makes it possible to effectively identify focal genomic aberrations in broadly aberrant regions. The proposed

method integrates multiple cancer samples so that it enables the detection of the consistent aberrations across multiple samples. We then apply this method to glioblastoma multiforme and lung cancer data sets from the SNP microarray platform. Through this process, we confirm the ability to detect previously known cancer related genes from both cancer types with high accuracy. Also, the application of this approach to a lung cancer data set identifies focal amplification regions that contain known oncogenes, though these regions are not reported using a recent CNAs detecting algorithm GISTIC: SMAD7 (chr18q21.1) and FGF10 (chr5p12). Our results suggest that WIFA can be used to reveal cancer related genes in various cancer data sets.

The talk is based on our paper:
Hur and Lee, Wavelet-based identification of DNA focal genomic aberrations from single nucleotide polymorphism arrays. BMC Bioinformatics 2011, 12:146

15:15-15:40 *Exploiting a Reduced Set of Weighted Average Features to Improve Prediction of DNA-binding Residues from 3D Structures*

Juan Liu
School of Computer, Wuhan University, China

Abstract: Predicting DNA-binding residues from a protein three-dimensional structure is a key task of computational structural proteomics. In the present study, based on the machine learning technology, we aim to explore a reduced set of weighted average features for improving prediction of DNA-binding residues on protein surfaces. Via constructing the spatial environment around a DNA-binding residue, a novel weighting factor is firstly proposed to quantify the distance-dependent contribution of each neighboring residue in determining the location of a binding residue. Then, a weighted average scheme is introduced to represent the surface patch of the considering residue. Finally, the classifier is trained on the reduced set of these weighted average features, consisting of evolutionary profile, interface propensity, betweenness centrality and solvent surface area of side chain. Experimental results on 5-fold cross validation and independent tests indicate that the new feature set are effective to describe DNA-binding residues and our approach has significant better performance than two previous methods. Furthermore, a brief case study suggests that the weighted average features are powerful for identifying DNA-binding residues and are promising for further study of protein structure-function relationship.

The talk is based on our paper:
Xiong Y, Xia J, Zhang W, Liu J. Exploiting a reduced set of weighted average features to improve prediction of DNA-binding residues from 3D structures. PLoS One. 2011;6(12):e28440.

15:40-16:05 *PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites*

Jiangning Song
Department of Biochemistry and Molecular Biology, Monash University Australia and Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, China

Abstract: The ability to catalytically cleave protein substrates after synthesis is fundamental for all forms of life. Accordingly, site-specific proteolysis is one of the most important post-translational modifications. The key to our understanding of the physiological role of a protease is to identify its natural substrate(s). Knowledge of the substrate specificity of a protease can dramatically improve our ability to predict its target protein substrates, but this information must be utilized in an effective manner by in silico approaches in order to efficiently identify protein substrates. To address this problem, we present PROSPER, an integrated feature-based server for in silico identification of protease substrates and their cleavage sites for twenty-four different protease families. PROSPER utilizes established specificity information for these proteases (derived from the MEROPS database) with a machine learning approach to predict protease cleavage sites by using different, but complementary sequence and structure characteristics. Features used by PROSPER include local amino acid sequence profile, predicted secondary structure, solvent accessibility and predicted native disorder. Thus, for proteases with known amino acid specificity, PROSPER provides a convenient, pre-prepared tool for use in identifying protein substrates for the enzymes. Systematic prediction analysis for the twenty-four proteases thus far included in the database revealed that the features we have included in the tool strongly improve performance in terms of cleavage site prediction, as evidenced by their contribution to the performance improvement. In comparison with two state-of-the-art prediction tools, PoPS and SitePrediction, PROSPER achieves a greater accuracy and coverage. To our knowledge, PROSPER is the first comprehensive server capable of predicting cleavage sites of multiple proteases within a single substrate sequence using machine learning techniques.

The talk is based on our paper:
Song J, Tan H, Perry AJ, Akutsu T, Webb GI, Pike RN & Whisstock JC. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. Submitted for publication

14:00-16:05 ISB Regular Session B1 (Conference Room Pingpong at Xi'an Jianguo Hotel)

Topic: Network Biology
Chair: Masahiro Okamoto

14:00-14:25 *Hierarchical Modular Structure in Gene Co-expression Networks*

Shuqin Zhang

Center for Computational Systems Biology, School of Mathematical Sciences, Fudan University, Shanghai, 200433, China

Paper ID: 33

Abstract: Network module (community) structure has been a hot research topic in recent years. Many methods have been proposed for module detection and identification. Hierarchical structure of modules is shown to exist in different kinds of biological networks. Compared to the module identification methods, less research is done on the hierarchial structure of modules. In this paper, we propose a method for constructing the hierarchical modular structure in networks based on the extended random graph model. Statistical tests are applied to test the hierarchial relations between different modules. We give both artificial networks and real data examples to illustrate the performance of our approach. Application of the proposed method to yeast gene co-expression network shows that it does have a hierarchical modular structure with the modules on different levels corresponding to different gene functions.

14:25-14:50 *A Gaussian Graphical Model for Identifying Significantly Responsive Regulatory Networks from Time Series Gene Expression Data*

Zhi-Ping Liu, **Wanwei Zhang**, Katsuhisa Horimoto, Luonan Chen

Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China

Paper ID: 39

Abstract: With rapid accumulation of functional relationships between biological molecules, knowledge-based networks have been constructed and stocked in many databases. These networks provide the curated and comprehensive information for the functional linkages among genes and proteins, while their activities are highly related with specific phenotypes and conditions. To evaluate a knowledge-based network in a specific condition, measuring the consistency between its structure and the conditionally specific gene expression profiling data is an important criterion. In this work, we propose a Gaussian graphical model to evaluate the documented regulatory networks by the consistency between network architectures and time-series gene expression profiles. By developing a dynamical Bayesian network model, we derive a new method to evaluate gene regulatory networks in both simulated and true time series microarray data. The regulatory networks are evaluated by matching a network structure and gene expressions, which are achieved by randomly rewiring the regulatory structures. To demonstrate the effectiveness of our method, we identify the significant regulatory networks in response to the time series gene expression of circadian rhythm. Moreover, the knowledge-based networks are screened and ranked by their consistencies of structures based on dynamical gene expressions.

14:50-15:15 *A Seed-based Approach to Identify Risk Disease Sub-networks in Human Lung Cancer*

Yi-Bin Wang, Yong-Mei Cheng, Shao-Wu Zhang

School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China

Paper ID: 35

Abstract: Lung cancer is the leading cause of cancer deaths worldwide. The identification of lung cancer risk disease sub-networks not only helps to understand lung cancer mechanism better, but also provide the potential benefits for the early diagnosis and lead to important applications such as drug targeting. Although some researches are devoted to investigating the carcinogenic process of lung cancer, these approaches have still some limitation. In this paper, the differentially expressed genes are scored and ranked in according to the method of augmented fuzzy measure similarity for obtaining the seed genes. Then, the model of random walk with restarts is used to identify risk disease sub-networks in the PPI network. At last 37 risk disease sub-networks are exploited from the PPI network, which play an important potential role in the carcinogenic process of the lung cancer disease. In terms of the proof and comments in the existing literatures, the identified results show that the proposed method works well in identifying the significant lung cancer risk disease sub-networks, and it is also suitable to recognize other complex risk disease sub-networks.

15:15-15:40 *Pigmented Network Structure Detection Using Semi-Smart Adaptive Filters*

Leszek A. Nowak, Maciej J. Ogorzalek, Marcin P. Pawlowski

Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, Krak ów,

Poland
Paper ID: 74
Abstract: This paper demonstrates a method for detecting pigment based dermatoscopic structure called pigment network. This structure is used in dermatoscopy as one of the criteria in clinical evaluation of pigmented skin lesions and can indicate if a lesion is of malignant nature. For detection process we have developed an adaptive filter, inspired by Swarm Intelligence (SI) optimization algorithms. The introduced filtering method is applied in a non-linear manner, to processed dermatoscopic image of a skin lesion. The non-linear approach derives from SI algorithms, and allows selective image filtering. In the beginning of filtration process, the filters (agents) are randomly applied to sections of the image, where each of them adapts its output based on the neighborhood surrounding it. Agents share its information with other agents that are located in immediate vicinity. This is a new approach to the problem of dermatoscopic structure detection, and it is highly flexible, as it can be applied to images without the need of previous pre-processing stage. This feature is highly desirable, mainly due to the fact that in most cases of computer aided diagnostic, input images need to be preprocessed (e.g.: brightness normalization, histogram equation, contrast enhancement, color normalization) and results of this can introduce unwanted artifacts, so this step need to be verified by human. Results of applying the introduced method can be used as one of the differential structures criteria for calculating the Total Dermatoscopy Score (TDS) of the ABCD rule.

15:40-16:05 *ppiPre - an R package for predicting protein-protein interactions*

Yue Deng, Lin Gao
Department of Computer Science, School of Software, Engineering, Xidian University, Xi'an, China
Paper ID: 79
Abstract: Since the existing experimental techniques for identifying protein-protein interactions (PPIs) are expensive and time-consuming, and the results are incomplete and/or noisy, developing computational methods for effectively predicting PPIs is of great importance. Therefore, we develop the R package ppiPre, which predicts PPIs using heterogeneous data sources, including Gene Ontology (GO) annotations, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations and topological properties of the PPI network. ppiPre supports up to 20 species and provides useful functions for predicting PPIs and calculating semantic and topological similarities between proteins. ppiPre is open source and freely available from http://cran.r-project.org/web/packages/ppiPre).

16:05-16:25 Coffee break

16:25-18:30 ISB Highlight Session H2 (Conference Room Peony at Xi'an Jianguo Hotel)

Topic: Proteomics
Chair: Ling-Yun Wu

16:25-16:50 *Systematic analysis of the Plk-mediated phosphoregulation in eukaryotes*

Yu Xue
Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, China
Abstract: Substantial evidence has confirmed that Polo-like kinases (Plks) play a crucial role in a variety of cellular processes via phosphorylation-mediated signaling transduction. Identification of Plk phospho-binding proteins and phosphorylation substrates is fundamental for elucidating the molecular mechanisms of Plks. Here, we present an integrative approach for the analysis of Plk-specific phospho-binding and phosphorylation sites (p-sites) in proteins. From the currently available phosphoproteomic data, we predicted tens of thousands of potential Plk phospho-binding and phosphorylation sites in eukaryotes, respectively. Furthermore, statistical analysis suggested that Plk phospho-binding proteins are more closely implicated in mitosis than their phosphorylation substrates. Additional computational analysis together with in vitro and in vivo experimental assays demonstrated that human Mis18B is a novel interacting partner of Plk1, while pT14 and pS48 of Mis18B were identified as phospho-binding sites. Taken together, this systematic analysis provides a global landscape of the complexity and diversity of potential Plk-mediated phosphoregulation, and the prediction results can be helpful for further experimental investigation.

The talk is based on our paper:
Liu Z, Ren J, Cao J, He J, Yao X, Jin C, Xue Y. (2012) Systematic analysis of the Plk-mediated phosphoregulation in eukaryotes. Briefings in Bioinformatics, in press.

16:50-17:15 *Systematic analysis and prediction of longevity genes in Caenorhabditis elegans*

Yan-Hui Li
Bioinformatics Group, Scientific Data Center, Computer Network Information Center, Chinese Academy of Sciences, China

Abstract: An important task of aging research is to find genes that regulate lifespan. However, identification of genes related to longevity (referred to as longevity genes hereafter) through web-lab experiments such as genetic screens is a tedious and labor-intensive activity. Developing an algorithm to predict longevity genes should facilitate aging research. In this paper, we systematically analyzed properties of longevity genes in Caenorhabditis elegans and found that, when compared to genes not yet known to be involved in longevity, known longevity genes display the following features: (i) longer genomic sequences and protein sequences, (ii) a stronger tendency to co-express with other genes during a transition from dauer state (an extremely long lifespan) to non-dauer state (a normal lifespan), (iii) significant enrichment in certain functions and RNAi phenotypes, (iv) higher sequence conservation, and (v) higher in several network topological features such as degrees in a functional interaction network. Based on these features, we developed an algorithm to predict longevity genes in C. elegans and obtained 243 novel longevity genes with a precision rate of 0.85. Some of the predicted genes have been validated by published articles or wet lab experiments. The contribution of each feature to the predicted results was also evaluated.

The talk is based on our paper:
Li YH, Dong MQ, Guo Z. Systematic analysis and prediction of longevity genes in Caenorhabditis elegans. Mech Ageing Dev. 2010 Nov-Dec;131(11-12):700-9.

17:15-17:40 *Unrestrictive Identification of Protein Post-translational Modifications from Tandem Mass Spectra*

Yan Fu
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China

Abstract: Most proteins in cells carry some posttranslational modifications (PTMs), and PTMs often play an essential role in protein functions and are involved in many pathological processes. Currently, several hundreds of types of PTMS have been known, but only a few of them are investigated in detail. Tandem mass spectrometry-based proteomics provides a powerful tool for large-scale analysis of PTMs. Computational analysis of the tandem mass spectra, especially those produced from peptides carrying potential PTMs, is a challenging problem in proteomics. Identification of modified proteins from tandem mass spectra is traditionally done in a restrictive manner, in which, only a small set of fixed PTM types are considered in data analysis. To more extensively detect PTMs in a protein sample, unrestrictive approaches to PTM identification have been proposed. Though promising, existing methods for unrestrictive identification of PTMs faces some difficult problems, such as low computational speed and high level of false matches. In this talk, I will present some of our recent research progresses on new algorithms and strategies for efficient and reliable unrestrictive identification of PTMs. Three algorithms for PTM discovery will be described, including fast discovery of abundant PTMs using peptide precursor information, open sequence database search, and open spectral library search. Factors influencing the false discovery rates (FDR) of PTM identifications will be discussed.

The talk is based on our papers:
[1] Ding Ye, Yan Fu, Ruixiang Sun, Haipeng Wang, Zuofei Yuan, Hao Chi and Simin He. Open MS/MS Spectral Library Search to Identify Unanticipated Post-Translational Modifications and Increase Spectral Identification Rate. Bioinformatics (ISMB2010), 26(12):i399-i406, 2010.
[2] Yan Fu, Liyun Xiu, Wei Jia, Ding Ye, Ruixiang Sun, Xiaohong Qian, Si-min He. DeltAMT: a statistical algorithm for fast detection of protein modifications from LC-MS/MS data. Molecular & Cellular Proteomics, 10(5):M110.000455, 2011.
[3] Yan Fu. Bayesian false discovery rates for post-translational modification proteomics. Statistics and Its Interface, 5:47–59, 2012.

17:40-18:05 *Comparative Analysis of Different Label-Free Mass Spectrometry Based Protein Abundance Estimates and Their Correlation with RNA-Seq Gene Expression Data*

Ning Kang
Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, China

Abstract: Complex chronic diseases are usually not caused by changes in a single causal gene but by an unbalanced regulating network resulting from the dysfunctions of multiple genes or their products. Therefore, network based systems approach can be helpful for the identification of candidate genes related to complex diseases and their relationships. The Axial spondyloarthropathy (SpA) is a group of chronic inflammatory joint diseases that mainly affects the spine and the sacroiliac joints, yet, the pathogenesis of SpA remains largely unknown. In this paper, we

conducted a networked systems study on the pathogenesis of SpA. We integrated data related to SpA, from the OMIM database, proteomics and microarray experiments of SpA, to prioritize SpA candidate disease genes in the context of human protein interactome. Based on the top ranked SpA related genes, we constructed a PPI network and identified potential pathways associated with SpA. The PPI network and pathways reflect the well-known knowledge of SpA, i.e., immune mediated inflammation, as well as imbalanced bone modeling caused new bone formation and bone loss. This study may facilitate our understanding of the SpA pathogenesis from the perspective of network systems.

This talk is based on our paper:
Ning K, Fermin D, Nesvizhskii AI. Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J Proteome Res. 2012 Apr 6;11(4):2261-71.

18:05-18:30 *Bridging modern and traditional Chinese medicines by knowledge-mining*

Xiaoqin Yang
National Scientific data sharing platform for population and health, China

Abstract: Unlike Western medicine, Traditional Chinese Medicine (TCM), which is based on the doctrine and empirical practices of systems science, not only uses simple yet meaningful symptoms or their combinations to describe diseases, but also provides personalized prescription of herbal medicine from a system perspective. However, the theory behind them is often obscure for the lack of scientific basis and molecule evidence as compared to modern medicine. TCMeSH, a component of VeryGene project (http://www.verygene.com), was developed as an attempt to fill this gap. A significant effort has been made to link TCM symptoms and standard disease terms by common genes using text mining. The herb-compound-gene relationships were also manually curated after data integrating and multi-level mapping. The current release carefully selected 4,989 gene-symptom relationships, covering 80 TCM symptoms. Meanwhile, over 6,000 interactions between human genes and 586 active compounds from more than 1,300 reputable Chinese herbs have been captured. Additional gene attributes (such as tissue specificity, biological pathways, Gene Ontology and mammalian phenotypes) were integrated. All these make TCMeSH not only a TCM genomics database, but also a discovery tool by generating novel inferences. Some inherently useful but hidden relations among genes, diseases, TCM symptoms can be inferred. Meanwhile, the wealth of herbal medicine–gene–disease information yields testable hypotheses for understanding the effects of TCM on human health.

This talk is based on our paper:
X. Yang, Y. Ye, G. Wang, H. Huang, D. Yu, and S. Liang, "VeryGene: linking tissue-specific genes to diseases, drugs and beyond for knowledge discovery," Physiological Genomics. 43. 457-460. 2011.

16:25-18:30 ISB Regular Session B2 (Conference Room Pingpong at Xi'an Jianguo Hotel)

Topic: Dynamics Systems Biology
Chair: Fangxiang Wu

16:25-16:50 *The influence of the basin structure of Boolean networks on their long range correlated dynamics*

Peng Xu , Xianghong Wang, **Wenbin Liu**
Department of Physics and Electronic information engineering, Wenzhou University, Wenzhou 325035, Zhejiang, China

Paper ID: 4

Abstract: It has been known for quite some time that the 1/ f dynamics play a vital role in living organisms. Recently we studied the long-range correlated dynamics of Boolean networks, and found that some networks could present the 1/ f dynamics while others couldn't. An important question is what kind of networks can generate such dynamics? In this paper, we investigate this issue based on the attractor structure of Boolean networks. We find that multiple attractor networks prefer to generate the 1/ f dynamics and systems with large basin entropy tend to sustain such dynamics in a wide noise range. Models for eight real genetic networks also partially support these observations.

16:50-17:15 *Functional tunability of biological circuits from tinkers*

Changhong Shi, Tianshou Zhou
School of Mathematics and Computational Sciences and Guangdong Province Key Laboratory of Computational Science, Sun Yat-Sen University, Guangzhou, P.R. China

Paper ID: 13

Abstract: In many complex regulatory networks with interlinked feedback loops, the simple core circuits are sufficient

to achieve specific biological functions of the entire networks, naturally raising a question: what is the role of the additional feedback loops. Here, by investigating auto-activation and activator-repressor circuits, two most common functional motifs in regulatory networks, we show that the toggle switch acts as a tinker to elaborate the dynamical behavior of both circuits. Specifically, the additional loop does not significantly affect the stable states of the auto-activation circuit but can tune the stimulation threshold for switch (i.e., the minimal stimulus required to switch the system from the low to the high state). For the activator- repressor circuit, the tinker can tune the stimulation threshold for oscillation (i.e., the bifurcation point to generate oscillations) as well as the oscillation frequency but does not change the oscillation amplitude. These detailed results not only provide guidelines to the engineering of both synthetic circuits but also imply a significant fact that additional loops of the core circuit in a complex network are not really redundant but play a role of tuning the network's function

17:15-17:40 *Escape from infinite adaptive peak*

Song Xu, Shuyun Jiao, Pengyao Jiang, Bo Yuan, Ping Ao
Department of Computer Science and Engineering, Shanghai Jiao Tong University, 200240, Shanghai, China

Paper ID: 65

Abstract: We study the transition time between different metastable states in the continuous Wright-Fisher (diffusion) model. We construct an adaptive landscape for describing the system both qualitatively and quantitatively. When strong genetic drift and weak mutation generate infinite adaptive peaks, we calculate the expected time to escape from such peak states. We find a new way to analytically approximate the escape time, which extends the application of Kramer's classical formulae to the cases of non-Gaussian equilibrium distribution and bridges previous results in two limits. Our adaptive landscape, compared to the classical fitness landscape or other scalar functions, is directly related to system's middle-and-long-term dynamics and is self-consistent in the whole parameter space. Our work provides a complete description for the bi-stabilities in the present model.

17:40-18:05 *A Stable Simplification of a Fas-signaling Pathway Model for Apoptosis*

Ya-Jing Huang, Wen-An Yong
Zhou Pei-Yuan Center for Appl. Math.Tsinghua Univ., Beijing 100084, China

Paper ID: 34

Abstract: Apoptosis is important for maintaining normal embryonic development, tissue homeostasis and normal immunesystem operation in multicellular organisms. Its malfunction may result in serious diseases such as cancer, autoimmunity, and neurodegeneration. In apoptosis, tens of species are present in many biochemical reactions with times scales of widely differing orders of magnitude. According to the law of mass action, apoptosis is usually described with a large and stiff system of ODEs (ordinary differential equations). The goal of this work is to derive a simple system of ODEs by using the classical PEA (partial equilibrium approximation) method. For this purpose, we firstly justify the mathematical correctness of the PEA in a quite general framework. On the basis of this result, we simplify the Fas-signaling pathway model proposed by Hua et al. (2005) by assuming the fastness of several reversible reactions. Numerical simulations and sensitivity analysis show that our simplification model is reliable.

18:05-18:30 *Switch-Like Regulation of Signal Transduction by Small RNA-mediated Quorum Sensing*

Xi Liu, Peipei Zhou, Ruiqi Wang
Institute of Systems Biology, Shanghai University, Shanghai 200444, China

Paper ID: 43

Abstract: Quorum sensing (QS) is a mechanism by which bacteria produce, release, and then detect and respond to biosignals called autoinducers (AIs). There are multiple feedback loops in the QS system of Vibrio harveyi. However, how these feedback loops function to control signal processing remains unclear. In this paper, we present a computational model for switch-like regulation of signal transduction by small regulatory RNAmediated QS based on intertwined network involving AIs, LuxO, LuxU, Qrr sRNAs, and LuxR. In agreement with experimental observations, the model suggests that different feedbacks play critical roles in the switch-like regulation. Our results reveal that Vibrio harveyi uses multiple feedbacks to precisely control signal transduction.

18:30-19:30 Welcome Reception (Conference room Lily in Xi'an Jianguo Hotel)

19:00-21:30 Board member Meeting for Computational Systems Biology

Society of ORSC (Conference Room (To Be Determined) in Xi'an Jianguo Hotel)

August 19 (Sunday) Technical sessions

8:00-10:05 ISB Highlight Session H3 (Conference Room Peony at Xi'an Jianguo Hotel)

Topic: Bioinformatics

Chair: Junwen John Wang

08:00-08:25 *Network-based approach for gene set analysis*

Weidong Tian

Institute of Biostatistics, School of Life Sciences, Fudan University, China

Abstract: Gene set analysis (GSA) is widely used to identify whether certain biological functions are enriched within the genes inside a gene list, e.g., differentially expressed genes, which can help to elucidate the molecular mechanism behind cell development, disease, etc.. Many statistical methods have been developed for GSA. However, most those methods treated genes inside a give gene set equally, while ignoring the inherent functional non-equivalence among genes in a given biological pathway. To account for the functional non-equivalence among a set of genes, we introduce a network-based approach for gene set analysis, which up-weights genes with functions more relevant to the pathway of interest. The genes are weighted according to its degree distribution inside a genome-scale functional association network. Based on a number of data analysis, we have demonstrated the power and reproducibility of our proposed method over the traditional unweighted approaches, making it a useful new method for gene set analysis.

This talk is based on our paper:
Fang Z, Tian W, Ji H. A network-based gene-weighting approach for pathway analysis. Cell Res. 2012 Mar;22(3):565-80. doi: 10.1038/cr.2011.149.

08:25-08:50 *Evolution and genetic variation of microRNA mediated gene regulation in humans*

Zhaolei Zhang

Banting and Best Department of Medical Research, The Donnelly Centre, University of Toronto, Toronto, Canada

Abstract: MicroRNA (miRNA) mediated gene regulation is of critical functional importance in animals and is often thought to be largely constrained during evolution. Here we show that a number of miRNA binding sites display high level of population differentiation in humans and thus are likely targets of local adaptation. In a subset we demonstrate that allelic differences modulate miRNA regulation in mammalian cells, including an interaction between miR-155 and TYRP1, a melanosomal enzyme associated with human pigmentary differences. We identify alternate alleles of TYRP1 that induce or disrupt miR-155 regulation and demonstrate that these alleles are selected with different modes among human populations, to optimize the protein abundance in response to different level of UV radiation. Our findings illustrate the evolutionary plasticity of the microRNA regulatory network in recent human evolution.

This talk is based on our paper:
Li Ji(e)., Liu Y.(e), Xin X., Kim T., Cabeza, EA, Ren J., Nielsen R., Wrana JL*, Zhang Z* "Evidence for positive selection on a number of microRNA regulatory interactions during recent human evolution" PLoS Genet. 2012 Mar;8(3):e1002578

08:50-09:15 *Discovery of multi-dimensional modules by integrative analysis of cancer genomic data*

Shihua Zhang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Recent technology has made it possible to simultaneously perform multi-platform genomic profiling (e.g., DNA methylation, and gene expression) of biological samples, resulting in so-called ``multi-dimensional genomic data". Such data provide unique opportunities to study the coordination between regulatory mechanisms on multiple levels. However, integrative analysis of multi-dimensional genomics data for the discovery of combinatorial patterns is currently lacking. Here, we adopt a joint matrix factorization technique to address this challenge. This method projects multiple types of genomic data onto a common coordinate system, in which heterogeneous variables weighted highly in the same projected direction form a multi-dimensional module. Genomic variables in such modules are characterized

by significant correlations and likely functional associations. We applied this method to the DNA methylation, gene expression, and microRNA expression data of 385 ovarian cancer samples from the TCGA project. These multi-dimensional modules revealed perturbed pathways that would have been overlooked with only a single type of data, uncovered associations between different layers of cellular activities, and allowed the identification of clinically distinct patient subgroups. Our study provides an useful protocol for uncovering hidden patterns and their biological implications in multi-dimensional ``omic'' data.

This talk is based on our papers:
[1] S. Zhang, C.-C. Liu, W. Li, H. Shen, P.W. Laird, X.J. Zhou. Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. Nucleic Acids Research, 2012, doi: 10.1093/nar/gks725.
[2] W. Li*, S. Zhang*, C.-C. Liu, X.J. Zhou. Identifying multi-layer gene regulatory modules from multi-dimensional genomic data. Bioinformatics, 2012, doi:10.1093/bioinformatics/bts476.
[3] S. Zhang, Q. Li, J. Liu, X.J. Zhou. A novel computational framework for simultaneous integration of multiple functional genomic data to identify microRNA-gene regulatory modules. Bioinformatics (ISMB2011), 2011, 27:i401-i409.

09:15-09:40 *MetaBinG: Using GPUs to Accelerate Metagenomic Sequence Classification*

Chaochun Wei
School of Life Sciences and Biotechnology, Shanghai Jiao Tong University

Abstract: Metagenomic sequence classification is a procedure to assign sequences to their source genomes. It is one of the important steps for metagenomic sequence data analysis. Although many methods exist, classification of high-throughput metagenomic sequence data in a limited time is still a challenge. We present here an ultra-fast metagenomic sequence classification system (MetaBinG) using graphic processing units (GPUs). The accuracy of MetaBinG is comparable to the best existing systems and it can classify a million of 454 reads within five minutes, which is more than 2 orders of magnitude faster than existing systems. MetaBinG is publicly available at http://cbb.sjtu.edu.cn/,ccwei/pub/software/MetaBinG/MetaBinG.php.

This talk is based on our paper:
Jia P, Xuan L, Liu L, Wei C (2011) MetaBinG: Using GPUs to Accelerate Metagenomic Sequence Classification. PLoS ONE 6(11): e25353. doi:10.1371/journal.pone.0025353

09:40-10:05 *Meta-Storms: Efficient Search for Similar Microbial Communities Based on a Novel Indexing Scheme and Similarity Score for Metagenomic Data*

Xiaoquan Su
Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, China

Abstract: It has long been intriguing scientists to effectively compare different microbial communities (also referred as “metagenomic samples” here) in a large scale: given a set of unknown samples, find similar metagenomic samples from a large repository, and examine how similar these samples are. With the current metagenomic samples accumulated, it is possible to build a database of metagenomic samples of interests. Any metagenomic samples could then be searched against this database to find the most similar metagenomic sample(s). However, on one hand, current databases with a large number of metagenomic samples mostly serve as data repositories that offer few functionalities for analysis; and on the other hand, methods to measure the similarity of metagenomic data work well only for small set of samples by pair-wise comparison. It is not yet clear how to efficiently search for metagenomic samples against a large metagenomic database. In this study, we have proposed a novel method, Meta-Storms, that could systematically and efficiently organize and search metagenomic data. It includes the following components: (i) creating a database of metagenomic samples based on their taxonomical annotations, (ii) efficient indexing of samples in the database based on a hierarchical taxonomy indexing strategy, (iii) searching for a metagenomic sample against the database by a fast scoring function based on quantitative phylogeny, and (iv) managing database by index export, index import, data insertion, data deletion and data-base merging. We have collected more than 1300 metagenomic data from the public domain and in-house facilities, and tested the Meta-Storms method on these datasets. Our experimental results show that Meta-Storms is capable of database creation and effective searching for a large number of metagenomic samples, and it could achieve similar accuracies compared to the current popular signifi-cance-testing based methods. Meta-Storms method would serve as a suitable database management and search system to quickly identify similar metagenomic samples from a large pool of samples.

This talk is based on our paper:
Su X, Xu J, Ning K. Meta-Storms: Efficient Search for Similar Microbial Communities Based on a Novel Indexing

Scheme and Similarity Score for Metagenomic Data. Bioinformatics. 2012 Jul 26.

08:00-10:05 ISB Session B3 (Conference Room Pingpong at Xi'an Jianguo Hotel)

Topic: Network Biology

Chair: Raul Rabadan

08:00-08:25 *Construction and Analysis of Genome-wide SNP Networks*

Yang Liu, Jin Zhou, Zhiping Liu, Luonan Chen and **Michael K. Ng**
Centre for Mathematical Imaging and Vision, Department of Mathematics, Hong Kong Baptist University, Hong Kong

Paper ID: 78

Abstract: The study of gene regulatory network (GRN) and protein protein interaction network (PPI) is believed to be fundamental to the understanding of molecular processes and functions in system biology and therefore, attracted more and more attentions in past few years. However, there is little focus about network construction in single nucleotide polymorphism (SNP) level, which may provide a direct insight into mutations among individuals, potentially leading to new pathogenesis discovery and diagnostics. In this paper, we present a novel method to mine, model and evaluate a SNP sub-network from SNP-SNP interactions. Specifically, based on logistic regression between two SNPs, we first construct a genome-wide SNP-SNP interaction network. Then by using gene information, selected SNP seeds are employed to detect SNP sub-networks with a maximal modularity. Finally to identify functional role of each SNP sub-network, its gene association network is constructed and their functional similarity values are calculated to show the biological relevance. Results show that our method is effective in SNP sub-network extraction and gene function prediction.

08:25-08:50 *Effective Clustering of MicroRNA Sequences by N-grams and Feature Weighting*

Yuan Yi, Jihong Guan, **Shuigeng Zhou**
Shanghai Key Lab of Intelligent Information Processing, and School of Computer Science, Fudan University, Shanghai 200433, China

Paper ID: 51

Abstract: MicroRNA (miRNA in short) is a kind of small RNAs that acts as an important post-transcriptional regulator with the Argonaute family of proteins to regulate target mRNAs in animals and plants etc. Since its first recognition as a distinct class of small RNA molecules in the early 1990s, tens of thousands of miRNAs have been identified experimentally or computationally. Currently, the focus of miRNAs study is on single-miRNA functions that usually result in gene silencing and repression. With the rapid increase of miRNAs, biologists have manually organized these miRNAs into biologically meaningful families to facilitate further study. As the members in the same family tend to share similar biochemical functions, a high quality family organization will shed lights on the functions of unknown miRNAs. However, manually grouping large amounts of miRNAs is not only time-consuming but also expensive. In this paper, we employ a clustering method with N-grams and feature weighting to automatically group miRNAs into separate clusters (families). Our method is evaluated with datasets constructed from the online miRNA database miRBase. Experimental results show that the clustering method can successfully distinguishes most miRNA families, and outperforms the traditional K-means clustering algorithm and the average-link clustering approach.

08:50-09:15 *Network Clustering along Diabetes Progression in Three Tissues of Goto-Kakizaki Rats*

Xinrong Zhou, Shigeru Saito, **Katsuhisa Horimoto**, Luonan Chen, Huarong Zhou
SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Chinese Academy of Sciences, Shanghai 200233, China

Paper ID: 25

Abstract: We investigated the macroscopic changes in the regulatory coordination of diabetes progression during three periods in three tissues, adipose, liver and muscle, of Goto-Kakizaki (GK) rats. For this purpose, we performed network clustering by the Newman algorithm for the regulatory networks inferred by a modified path consistency algorithm, and investigated the biological functions of each cluster by an enrichment analysis of the constituent genes. We then compared the network clusters characterized by biological functions with the diabetes progression of GK rats in each of the three tissues. The network structure, the number of clusters, and the number of clusters characterized by biological functions during the three periods showed similar patterns in the three tissues. In contrast, further scrutiny of the biological functions at coordinated clusters revealed characteristic differences between the three tissues along the diabetes progression. In particular, the hypothetical roles of each tissue emerged: adipose and liver function at the cellular and molecular levels at the early stage, respectively, and all three tissues are responsible for diabetes

progression, under the control of various transcriptional regulators.

09:15-09:40 *Network Kernel SVM for Microarray Classification and Gene Sets Selection*

Bing Yang, Junyan Tan, Naiyang Deng, Ling Jing
Department of Applied Mathematics, College of Science, China Agricultural University, 100083, Beijing, P.R. China

Paper ID: 29

Abstract: The importance of network-based approach to identifying biological markers has been increasingly recognized. Lots of papers indicated that genes in a network tend to function together in biological processes, so taking full advantage of the biological observation can improve the performance of microarray classification. However, lots of SVM methods don't consider this situation during their classifier building. The main idea of this paper intends to embed the information of gene networks into a new SVM learning framework. Based on a new regularization, we propose a novel method, Network Kernel SVM (NK-SVM), for binary classification problem and gene sets selection. By constructing some special kernel matrixes from the prior information of gene network, the new NK-SVM method makes the genes in the same set to be selected (or eliminated) together. The numerical experiments on a real microarray application show that the proposed method tends to provide a better performance than other methods on gene sets selection.

09:40-10:05 *cGRNexp: a Web Platform for Building Combinatorial Gene Regulation Networks based on user-uploaded gene expression datasets*

Huayong Xu, **Hui Yu**, Kang Tu, Qianqian Shi, Chaochun Wei,Yuanyuan Li, Yixue Li
School of Life Sciences and Biotechnology, Shanghai Jiao Tong University,100 Dongchuan Road, Shanhgai 200240, P.R.China

Paper ID: 32

Abstract: While we witness rapid progresses in development of methodologies/algorithms for constructing and analyzing the combinatorial regulation network which includes both TF regulators and miRNA regulators, we find a lack of tools or servers available for facilitating related works. A web service is especially needed that allows user to upload their own expression datasets and mine the combinatorial gene regulatory networks regarding the particular experimental context. Herein we report cGRNexp, a web platform for building combinatorial gene regulation networks based on user-uploaded gene expression datasets. In cGRNexp, we deposit three types of sequence-matching-based regulatory relationships and implement two functional modules for processing microRNAperturbed gene expression datasets and parallel miRNA/mRNA expression datasets. With the microarrays and next-generation sequencing platforms being increasingly accessible, a large amount of miRNA or mRNA expression datasets will be attainable in the near future, and thus, our web platform cGRNexp will be very useful for helping people mine the conditional combinatorial regulatory networks from their own expression datasets. cGRNexp is accessible at http://www.scbit.org/cgrnexp/.

10:05-10:25 Coffee break

10:25-12:30 ISB Highlight Session H4 (Conference Room Peony at Xi'an Jianguo Hotel)

Topic: Bioinformatics

Chair: Yu Xue

10:25-10:50 *Modeling and comparing the organization of circular genomes*

Grace S. Shieh
Institute of Statistical Science, Academia Sinica, Taiwan

Abstract: Most prokaryotic genomes are circular with a single chromosome (called circular genomes), which consist of bacteria and archaea. Orthologous genes (abbreviated as orthologs) are genes directly evolved from an ancestor gene, and can be traced through different species in evolution. Shared orthologs between bacterial genomes have been used to measure their genome evolution. Here, organization of circular genomes is analyzed via distributions of shared orthologs between genomes. However, these distributions are often asymmetric and bimodal; to date, there is no joint

distribution to model such data. This motivated us to develop a family of bivariate distributions with generalized von Mises marginals (BGVM) and its statistical inference.

A new measure based on circular grade correlation and the fraction of shared orthologs is proposed for association between circular genomes, and a visualization tool developed to depict genome structure similarity. The proposed procedures are applied to eight pairs of prokaryotes separated from domain down to species, and 13 mycoplasma bacteria that are mammalian pathogens belonging to the same genus. We close with remarks on further applications to many features of genomic organization, e.g. shared transcription factor binding sites, between any pair of circular genomes. Thus, the proposed procedures may be applied to identifying conserved chromosome backbones, among others, for genome construction in synthetic biology. All codes of the BGVM procedures and 1000+ prokaryotic genomes are available at http://www.stat.sinica.edu.tw/~gshieh/bgvm.htm.

This talk is based on our paper:
Shieh GS, Zheng S, Johnson RA, Chang YF, Shimizu K, Wang CC, Tang SL. Modeling and comparing the organization of circular genomes. Bioinformatics. 2011 Apr 1;27(7):912-8.

10:50-11:15 *Efficient Exponential Time Algorithms for Edit Distance between Unordered Trees*

Takeyuki Tamura

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Japan

Abstract: Edit distance of trees is one of the well-studied computational problems on tree structured data, which has applications in computational biology, for example, comparison of glycan structures. This talk presents efficient exponential time algorithms for the unordered tree edit distance problem, which is known to be NP-hard. For a general case, an $O(1.26^{n_1+n_2})$ time algorithm is presented, where n_1 and n_2 are the numbers of nodes in two input trees. This algorithm is obtained by a combination of dynamic programming, exhaustive search, and maximum weighted bipartite matching. For bounded degree trees over a fixed alphabet, it is shown that the problem can be solved in $O((1+\epsilon)^{n_1+n_2})$ time for any fixed $\epsilon > 0$. This result is achieved by avoiding duplicate calculations for identical subsets of small subtrees.

The talk is based on our paper on Combinatorial Pattern Matching 2012 (CPM2012).

11:15-11:40 *GWASdb: a database for human genetic variants identified by genome-wide association studies*

Mulin Jun Li

Department of Biochemistry & Centre for Genomic Sciences, The University of Hong Kong, China

Abstract: Recent advances in genome-wide association studies (GWAS) have enabled us to identify thousands of genetic variants (GVs) that are associated with human diseases. As next-generation sequencing technologies become less expensive, more GVs will be discovered in the near future. Existing databases, such as NHGRI GWAS Catalog, collect GVs with only genome-wide level significance. However, many true disease susceptibility loci have relatively moderate P values and are not included in these databases. We have developed GWASdb that contains 20 times more data than the GWAS Catalog and includes less significant GVs (P<1.010-3) manually curated from the literature. In addition, GWASdb provides comprehensive functional annotations for each GV, including genomic mapping information, regulatory effects (transcription factor binding sites, microRNA target sites and splicing sites), amino acid substitutions, evolution, gene expression and disease associations. Furthermore, GWASdb classifies these GVs according to diseases using Disease-Ontology Lite and Human Phenotype Ontology. It can conduct pathway enrichment and PPI network association analysis for these diseases. GWASdb provides an intuitive, multifunctional database for biologists and clinicians to explore GVs and their functional inferences. It is freely available at http://jjwanglab.org/gwasdb and will be updated frequently.

This talk is based on our paper:
Li MJ, Wang P, Liu X, Lim EL, Wang Z, Yeager M, Wong MP, Sham PC, Chanock SJ, Wang J. GWASdb: a database for human genetic variants identified by genome-wide association studies. Nucleic Acids Res. 2012 Jan;40(Database issue):D1047-54.

11:40-12:05 *Efficient methods for identifying mutated driver pathways in cancer*

Junfei Zhao

Academy of Mathematics & Systems Science, Chinese Academy of Sciences, China

Abstract: **Motivation:** The first step for clinical diagnostics, prognostics, and targeted therapeutics of cancer is to

comprehensively understand its molecular mechanisms. Large-scale cancer genomics projects are providing a large volume of data about genomic, epigenomic, and gene expression aberrations in multiple cancer types. One of the remaining challenges is to identify driver mutations, driver genes and driver pathways promoting cancer proliferation and filter out the unfunctional and passenger ones.
Results: In this study, we propose two methods to solve the so-called Maximum Weight Submatrix problem which is designed to de novo identify mutated driver pathways from mutation data in cancer. The first one is an exact method which can be helpful for assessing other approximate or/and heuristic algorithms. The second one is a stochastic and flexible method which can be employed to incorporate other types of information to improve the first method. Particularly, we propose an integrative model to combine mutation and expression data. We first apply our methods onto simulated data to show their efficiency. We further apply the proposed methods onto several real biological data sets such as the mutation profiles of 74 head and neck squamous cell carcinomas samples, 90 glioblastoma tumor samples and 313 ovarian carcinoma samples. The gene expression profiles were also considered for the later two data. The results show that our integrative model can identify more biologically relevant gene sets. We have implemented all these methods and made a package called MDPFinder (Mutated Driver Pathway Finder) which can be easily used for other researchers. Availability: A matlab package of MDPFinder is available at http://zhangroup.aporc.org/ShiHuaZhang. Contact: zsh@amss.ac.cn

This talk is based on our paper:
J. Zhao, S. Zhang*, L.-Y. Wu, X.-S Zhang. Efficient methods for identifying mutated driver pathways in cancer. Bioinformatics (minor revision), 2012.

12:05-12:30 *Using Graphical Adaptive Lasso approach to construct TF and miRNA's combinatorial regulatory network in breast cancer*

Naifang Su
School of Mathematical Sciences, Peking University, Beijing 100871, P.R.China

Abstract: Discovering the regulation of cancer related gene is of great importance in cancer biology. Transcription factors (TFs) and microRNAs (miRNAs) are two kinds of crucial regulators in gene expression, and they compose a combinatorial regulatory network with their target genes. Revealing the structure of this network could improve our understanding of the mechanisms underlying gene regulation, and further explore the molecular pathway in cancer. Here we propose a novel approach GALASSO (Graphical Adaptive lasso) to construct the regulatory network. GALASSO use a Gaussian graphical model with adaptive lasso penalties to integrate the sequence information as well as gene expression profiles. It is a comprehensive method for systematically study gene regulation. The simulation study and the experimental profiles verify the accuracy of our network. Meanwhile, the co-expression patterns and function annotations of target genes reveal their role in cancer. We further investigate the structure of the regulatory network, and discover some network motifs including feed forward loop. In addition, we introduce a quantity to measure miRNA and TF's co-regulatory effect, and explore the mechanism of every network patterns. Two types of feed forward loops have been distinguished. Moreover, we reveal the underlying mechanism of miR-155's role in breast cancer. Generally, the learned regulatory network promotes our understanding of cancer pathogenesis, and could help for future therapeutic studies.

10:25-12:30 ISB Regular Session B4 (Conference Room Pingpong at Xi'an Jianguo Hotel)

Topic: Next Generation Data Analysis
Chair: Paul Horton

10:25-10:50 *RNA-seq Coverage Effects on Biological Pathways and GO Tag Clouds*

Chien-Ming Chen, Tsan-Huang Shih, Tun-Wen Pai, Zhen-Long Liu, Margaret Dah-Tsyr Chang
Dept. of Computer Science and Engineering, 2Center of Excellence for Marine Bioenvironment and Biotechnology, National Taiwan Ocean University, Keelung, Taiwan

Paper ID: 58

Abstract: RNA-seq data analysis not only detects novel transcripts, promoters, and single nucleotide polymorphisms in a transcriptome scale, but also shows quantitative measurement of gene expression. In order to perform differential expression analysis for unraveling biological functions, we proposed a workflow which integrated annotations from KEGG biological pathways and Gene Ontology associations for manipulating multiple RNA-seq datasets. The developed system started from mapping short reads onto reference genes, and then performed normalization procedures on read coverage to evaluate and compare expression levels within various gene clusters. Different levels of gene expression were indicated by diverse color shades and graphically shown in designed temporal pathways. Besides,

representative GO terms associated with differentially expressed gene cluster were also visually displayed by a GO tag cloud representation. Three different public RNA-seq datasets were applied to demonstrate that the proposed workflow could provide effective and efficient analysis on differential gene expression for either cross-strain comparison or an identical sample sequenced at different time points.

10:50-11:15 *Statistical distribution of transcription (gene expression) in eukaryotic cells: insight from RNA-seq data.*

Hau Man Yeung, **Nelson Tang**
The Chinese University of Hong Kong

Abstract: With development of NGS in the past 3 years, the count of the number of transcripts provides a digital measure of gene expression in a sample. Studies have been shown that a gene count follows a Poisson or a Negative Binomial Distribution. However, there were few studies on the statistical distributions of gene expression in a genomic scale. It is obvious that all datasets are heavily skewed to the right. A generic form of distribution in a genomic scale for a sample can be used as a quality control in the analysis. We examine a dataset from Mastrokolias A et. al. [1], containing a set of RNA-seq results obtained from 6 healthy subjects. Each sample was divided into two processing methods, (1) without depletion of hemoglobin genes and (2) with depletion. As hemoglobin genes represent the most abundant transcripts in blood sample, it may mask expression of other genes. Firstly, we apply the both the data to estimate the corresponding Lorenz curve and find the Gini's coefficient for each sample. Dataset without depletion of hemoglobin, the mean is around 0.95 and variance in the order 10e-5, while the data with depletion, the mean is around 0.86 and same order of variance as above. This shows that a generic form of distribution does exist across samples. Secondly, we examined the goodness of fit for both dataset with original count and logarithmic counts by Kolmogorov–Smirnov Test to different distributions. Data are grouped according to their expression level ranking with different size. It is worth noting that at the lower end of the groups, all distributions are rejected. Both original and logarithmic count fit into Normal, and Uniform distributions in the same range for all samples. The least count that allow a fit with the above distributions for both datasets is about 150 counts. This suggests an analysis assumed with any distribution should take genes with more than 150 counts. Thirdly, we have examined the samples with depletion. This procedure allows the genes to be tagged more frequently at the lower end, which increase the number of genes to get a count over 150 and result in a distribution fit starting at a lower rank. This agrees with the previous point. Fourthly, we have examined the correlation for all genes across samples. By assuming all samples are healthy individuals, genes across samples and across depletion should have relative stable position. The spearman correlation is high (around 0.9, with p-value < 0.0005) among all samples.
Conclusion: Gaussian distribution can only be applied to genes with very similar expression level and does not apply to genes in the whole genome. There have been many conflicting results in previous study of gene biomarkers using microarrays. It may be accounted by the violation of the assumed Gaussian distribution and these gene markers were picked up due to statistical analysis using the wrong distribution model as assumption. This is particularly true for gene with low expression (counts < 150). It also provides insight for new approach of statistical analysis of gene expression. Here, we suggest that a kind distribution –free analysis is required.

11:15-11:40 *A machine learning framework of functional biomarker discovery for different microbial communities based on metagenomic data*

Wei Fang, Xingzhi Chang, Xiaoquan Su, Jian Xu, Deli Zhang, Kang Ning
Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, China

Paper ID: 30

Abstract: As more than 90% of microbial community could not be isolated and cultivated, the metagenomic methods have been commonly used to analyze the microbial community as a whole. With the fast acumination of metagenomic samples, it is now intriguing to find simple biomarkers, especially functional biomarkers, which could distinguish different metagenomic samples. Next-generation sequencing techniques have enabled the detection of very accurate gene-presence (abundance) values in metagenomic studies. And the presence/absence or different abundance values for a set of genes could be used as appropriate biomarker for identification of the corresponding microbial community's phenotype. However, it is not yet clear how to select such a set of genes (features), and how accurate would it be for such a set of selected genes on prediction of microbial community's phenotype. In this study, we have evaluated different machine learning methods, including feature selection methods and classification methods, for selection of biomarkers that could distinguish different samples. Then we proposed a machine learning framework, which could discover biomarkers for different microbial communities from the mining of metagenomic data. Given a set of features (genes) and their presence values in multiple samples, we first selected discriminative features as candidate by feature selection, and then selected the feature sets with low error rate and classification accuracies as biomarkers by classification method. We have selected whole genome sequencing data from simulation, public domain and in-house metagenomic data generation facilities. We tested the framework on prediction and evaluation of the biomarkers.

Results have shown that the framework could select functional biomarkers with very high accuracy. Therefore, this framework would be a suitable tool to discover functional biomarkers to distinguish different microbial communities.

11:40-12:05 *BAsplice: Bi-direction Alignment for detecting splice junctions*

Jingde Bu, Jiayan Wu, Meili Chen, Jingfa Xiao, Jun Yu, Xuebin Chi, Zhong Jin
Supercomputing Center, Computer Network Information Center, Chinese Academy of Sciences, Beijing, China

Paper ID: 60

Abstract: RNA-Seq is a revolutionary whole transcriptome shotgun sequencing technology performed by high-throughput sequencers, which provide more comprehensive information on differential expression of genes and benefit on novel splice variants identification. RNA-Seq reads is so short that it's a great challenge on mapping reads back to the reference effectively, especially when they span two or more exons. To improve the mapping efficiency, we introduce here a bidirection alignment tool – BAsplice, which use RNA-Seq data to detect splice junctions without any additional information. Compare with another splice junction mapping software, SOAPsplice, BAsplice performs better in call rate and running time, but a little worse in accuracy. BAsplice is a free open-source software written in C language. It is available at https://github.com/vlcc/basplice.

12:05-12:30 *Identifying Mutated Core Modules in Glioblastoma by Integrative Network Analysis*

Junhua Zhang, Shihua Zhang, Yong Wang, Junfei Zhao, Xiang-Sun Zhang
National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Paper ID: 72

Abstract: Glioblastoma multiforme (GBM) is the most common and aggressive type of brain tumor in humans. Distinguishing "driver" mutations from passively selected "passengers" is a central challenge in computational cancer biology. Because of mutational heterogeneity, analyses that extend beyond single genes are often restricted to examine known pathways and functional modules for enrichment of somatic mutations. In this paper we present a network-based method to identify mutated core modules for tumors without any prior information other than the data of somatic mutations and gene expressions from tumor patients. Firstly, two networks with weighted vertices and weighted edges are constructed by using the mutations and expressions, respectively. Then these two networks are combined to get an integrative network, for which an optimization model is used to identify the most coherent subnetworks. With the significance and exclusivity tests we get the core modules for tumors. By applying our method to The Cancer Genome Atlas (TCGA) GBM data, we obtained three core modules, which contain not only oncogenes and tumor suppressors that have been previously implicated in GBM pathogenesis (e.g., EGFR, TP53, PTEN, NF1 and RB1), but also some genes which have not or rarely been reported earlier in the context of glioblastoma multiforme (e.g., DST, PRAME and SYNE1). Thus, in addition to present generally applicable methodology, our findings provide several GBM candidate genes for further studies.

12:30-14:00 Lunch break (Buffet at Xi'an Jianguo Hotel)

14:00-16:05 ISB Highlight Session H5 (Conference Room Peony at Xi'an Jianguo Hotel)

Topic: Bioinformatics

Chair: Bairong Shen

14:00-14:25 *Understanding Virus Infection and Therapy based on Mathematical Modeling and Experiments*

Xiao Chen
School of Mathematics and Physics, University of Science and Technology Beijing, PR China

Abstract: **Background:** Mathematical modelling has made significant contributions to the fields of (anti-) HIV, HBV and /or HCV infections. The theoretical analysis of virus infection models based on experimental data may understand deeper the faction that govern infection disease progression. **Method:** Based on the hepatitis B virus (HBV) infection experimental data reported by M. A. Nowak et al., G.K. Lau et al., and Weiland et al.. We set up corresponding HBV

infection/anti-infection therapy models and numerically simulate the dynamics of HBV infections. **Results:** Each of the models has at least two constant solutions Q1 and Q2, representing virus free and persistent infection. Our study shows that if a patients basic virus infection number $R < 1$, then the patient will recover to Q1 state automatically or via treatment. Otherwise he/she will be in Q2 state. Numerical results can explain the corresponding experimental data. **Conclusions:** For suffering virus infecting population, they may be divided by four kinds: (1) Individuals have no any symptoms and recovered automatically even if infected with a large amount of virus if their $R < 1$. (2) Individuals have no any symptoms and become persistent infections even if infected with microscale virus if their $R > 1$. (3) Individuals have acute infection symptoms and recovered automatically even if infected with a large amount of virus if their $R < 1$. (4) Individuals have acute infection and become persistent infections even if infected with microscale virus if their $R > 1$. In cases (1) and (2), the basic virus infection number R depends on the death rates of hosts infection cells and virus, the rates of production of hosts infected cells and virus. In cases (3) and (4), the R depends on additional the rates of production and killing of hosts immune cells, the death rates of hosts immune cells, the languish rates of virus and infected cells caused by individual's immune ability or effect of anti-virus infection therapy. The research results suggest that if an anti-virus therapy has only blocked a patient's virus replications in vivo rather than activated the patient's immune function, the virus in the patient's body may reboot after stopping the treatment in the case that the patient's infected cells are not completely replaced by the normal ones. The reason why animal populations have hardly become extinct by various virus is that million years' biological evolutions make most of animal populations' basic virus infection number be less than one.

This talk is based on our papers:

[1] L. Min, Y. Su, and Y. Kuang, "Mathematical analysis of a basic model of virus infection with application to HBV infection," Rocky Mountain J. of Math., vol. 38, pp. 1573–1584, 2008.

[2] Y. Zheng, L. Min, and Y. J. et al., "Global stability of endemic equilibrium point of basic virus infection model with application to HBV infection," J. Systems Science and Complexity, vol. 23, pp. 1221–1230, 2010.

[3] Y. Zheng, L. Min, and X. Chen, "Dynamic analysis of HBV infection model with simulations for Anti-HBV infection therapy," in Proceedings of 2011 Int. Conf. on Intelligent Computation and Bio-Madical Instrumentation, vol. 2, Wuhan, China, 14-17 Dec. 2011, pp. 291–295.

[4] X. Chen, L. Min, and Y. K. et al., "Dynamics of acute hepatitis B virus infection in chimpanzees," Mathematics Computer and Simulation, submitted, 2011.

14:20-14:50 *CNetA: Network alignment by combining biological and topological features*

Qiang Huang, Ling-Yun Wu and Xiang-Sun Zhang.

Academy of Mathematics & Systems Science, Chinese Academy of Sciences, China

Paper ID: 54

Abstract: Due to the rapid progress of high-throughput techniques in past decade, a lot of biomolecular networks are constructed and collected in various databases. However, the biological functional annotations to networks do not keep up with the pace. Network alignment is a fundamental and important bioinformatics approach for predicting functional annotations and discovering conserved functional modules. Although many methods were developed to address the network alignment problem, it is not solved satisfactorily. In this paper, we propose a novel network alignment method called CNetA, which is based on the conditional random field model. The new method is compared with other four methods on three real protein-protein interaction (PPI) network pairs by using four structural and five biological criteria. Compared with structure-dominated methods, larger biological conserved subnetworks are found, while compared with the node-dominated methods, larger connected subnetworks are found. In a word, CNetA preferably balances the biological and topological similarities.

14:50-15:15 *FALCON ---- predicting protein structure prediction using fragment-HMM with an optimized energy function*

Dongbo Bu

Institute of Computing Technology , Chinese Academy of Sciences, China

Abstract: We designed a simple position-specific hidden Markov model to predict protein structure. Our new framework naturally repeats itself to converge to a final target, conglomerating fragment assembly, clustering, target selection, refinement, and consensus, all in one process. Our initial implementation of this theory converges to within 6 Å of the native structures for 100% of decoys on all six standard benchmark proteins used in ROSETTA (discussed by Simons and colleagues in a recent paper), which achieved only 14%–94% for the same data. The qualities of the best decoys and the final decoys our theory converges to are also notably better.

15:15-15:40 *A 3-Dimentional Multiscale Model to Simulate Tumor Progression in Response to Interactions between Cancer Stem Cells and Tumor Microenvironmental Factors*

Hua Tan, Fuhai Li, Jaykrishna Singh, Xiaofeng Xia, Derek Cridebring, Jian Yang, **Ming Zhan**, Stephen T.C. Wong, Jiguang Bao, Jinwen Ma

NCI Center for Modeling Cancer Development, Department of Systems Medicine and Bioengineering, The Methodist Hospital Research Institute, Weil Medical College of Cornell University Houston, Tx, U.S.A.

Paper ID: 71

Abstract: The recent discovery of cancer stem cells (CSCs), or tumor initiating cells (TICs), in a variety of cancers, including breast cancer, provides a key to understand the processes of tumor initiation, progression and recurrence. Here, we present a three-dimensional (3D) multiscale model of the CSC-initiated tumor growth, which takes into account essential microenvironmental (mE) factors (e.g. nutrients, extracellular matrix) and some important biological traits (e.g. angiogenesis, cell apoptosis, and necrosis) and addresses tumor growth from three different levels, i.e. molecular, cellular and tissue levels. At the molecular level, mathematical diffusion-reaction equations are used to understand the dynamics of mE factors. At the cellular level, a cellular automaton is designed to simulate the life cycle and behaviors of individual cells. At the tissue level, a computer graphics method is used to illustrate the geometry of the whole tumor. The simulation study based on the proposed model indicates that the content of CSCs in a tumor mass plays an essential role in driving tumor growth. The simulation also ighlights the significance of developing therapeutic agents that can deliver drug molecules into the interior of the tumor, where most of CSCs tend to reside. The simulation study on the breast cancer xenografts reveals that the mouse tumor initiated from a mixed population of human CSCs and other tumor cells show a faster growth rate, while a weaker proliferation and aggressiveness than that initiated from a pure human CSCs population. These simulation results are mostly consistent with our experimental observations. The mathematical model thus provides a new framework for the modeling and simulation studies of CSC-initiated cancer development.

15:40-16:05 *Detecting early-warnings signal for complex diseases by quantifying conditional network entropy*

Rui Liu, Kazuyuki Aihara, and Luonan Chen

Institute of Industrial Science, The University of Tokyo, Japan

Abstract: A novel concept called dynamical network biomarker (DNB) was proposed recently to study the sudden change during a biological process, e.g. the progression of complex diseases, which is resulted from a critical transition from one state to another, corresponding to a bifurcation of the dynamical system for the underlying organism. By exploiting rich information from time-course high-throughput data and based on DNB theory, in this paper we develop a new computational method to efficiently detect the critical transition for a dynamical network by quantifying a type of conditional network entropy, which could be used to measure the network robustness, and thus gain the insight into disease mechanisms at a systems level. We theoretically prove that the critical reduction of network robustness can detect the sudden change of a biological system, e.g., the deterioration of a complex disease. A numerical simulation is given to show the effectiveness.

14:00-16:05 ISB Regular Session B5 (Conference Room Pingpong at Xi'an Jianguo Hotel)

Topic: Bioinformatics

Chair: Zhaolei Zhang

14:00-14:25 *Comparative analysis of protein-coding genes and long non-coding RNAs of prostate cancer between Caucasian and Chinese populations*

Fu-Yan Hu, Xing-Ming Zhao, Nelson L.S. Tang, **Yan Zhang**, Luonan Chen

Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

Paper ID: 69

Abstract: Prostate cancer is one of the most important public health problems in developed countries. To date, a systematic understanding of the pathogenesis of prostate cancer is still lacking. In this work, we identified differentially expressed protein-coding genes and long non-coding RNAs (lncRNAs) between normal and cancer tissues based on a recent RNA-seq study from Caucasian population. We then investigated the relationship between differentially expressed genes and lncRNAs. Furthermore, based on a recently published prostate cancer study on Chinese population,

we identified differentially expressed genes between Caucasian and Chinese populations to investigate racial difference. Moreover, for the first time, we compared the correlation of lncRNA-gene across populations. In the end, a lot of differentially expressed genes and lncRNAs were identified. Our results revealed that most of the lncRNA-gene pairs were positively correlated especially for the lncRNA-host gene pairs, indicating the probable mechanism of lncRNA. And 320 genes were differentially expressed in prostate cancer across populations, which may help us to investigate the ethnic differences of prostate cancer. In addition, our results suggested that lncRNAs regulate genes in different manners across populations. Our findings may help understand molecular events underlying prostate cancer development.

14:25-14:50 *A Sequence-Segmented Method Applied to the Similarity Analysis of Proteins*

Fen Kong, Xu-ying Nan, **Ping-an He**, Qi Dai, Yu-hua Yao
College of Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China

Paper ID: 76

Abstract: A 2-D graphical representation of protein sequences based on two classifications of amino acids is outlined. The method of dividing a long sequence into k segments (SSM) is introduced, so protein graph is divided into k segments, geometrical center of the points for all protein curve segment is given as descriptors of protein sequences. It is not only useful for comparative study of proteins, but also for encoding innate information about the structure of proteins. Finally, a simple example is taken to highlight the behavior of the new descriptor on protein sequences taken from the 12 baculoviruse proteins.

14:50-15:15 *New encoding schemes for prediction of protein Phosphorylation sites*

Zimo Yin, Junyan Tan
College of Science, China Agricultural University, Beijing, China 100083

Paper ID: 12

Abstract: Protein phosphorylation is involved in most cellular functions. Because of the importance of protein phosphorylation,many methods are conducted to identify the phosphorylation sites. Experimental methods for identifying phosphorylation sites are not only costly but also time consuming. Hence, computational methods are highly desired. In this paper, three new encoding methods, BinCTF(Binary-conjoint triad feature), CTF2(new conjoint triad feature) and BinCTF2(Binary-new conjoint triad feature), which are the modification of Binary and CTF encoding, are developed. Then an ensemble support vector machine is applied to predict the phosphorylation sites related to serine (S), threonine (T) and tyrosine (Y) residues. The numerical results indicate that the performance of new method is better.

15:15-15:40 *Analysis of Morphological Evolution in a Long-term Experiment with {\it Escherichia coli*

Fangshu Cui, Bo Yuan
Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Paper ID: 52

Abstract: Great attentions are still paid to the morphological evolution, such as the waiting time to the morphological stability in constant environment, the contributions of different evolutionary forces to the morphological evolution and so on, despite considerable progress. To investigate these issues, some biologists seek to carry out evolution experiments owing to the incompleteness and uncontrollability of the fossil record and the natural populations. We analyze the morphology (cell size) evolution observed from a long-term evolution experiment with Escherichia coli by Lenski et al. and explore these questions more rigorously. We adopt a population genetics model, the Wright-Fisher model, to describe this morphological evolution and calculate the estimates of the waiting time until the ultimate stasis (near stasis) in morphology (cell size) in the long-term experiment by simulations. These calculations have been verified to be in good accordance with the experimental data, which demonstrates the effectiveness of our model. We have shown how the per-locus mutation rate, the average selection advantage permutation and the population size devote to the morphology (cell size) evolution. Our results indicate that the selective advantage plays a powerful effect on this morphological evolution. By comparison, the mutation rate and population size have a weaker influence.

15:40-16:05 *Predicting Protein-RNA Residue-base Contacts Using Two-dimensional Conditional Random Field*

Morihiro Hayashida, Mayumi Kamada, Jiangning Song, Tatsuya Akutsu
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji,

Kyoto, 611-0011, Japan
Paper ID: 41
Abstract: Understanding of interactions between proteins and RNAs is essential to reveal networks and functions of molecules in cellular systems. Many studies have been done for analyzing and investigating interactions between protein residues and RNA bases. For interactions between protein residues, it is supported that residues at interacting sites have co-evolved with the corresponding residues in the partner protein to keep the interactions between the proteins. In our previous work, on the basis of this idea, we calculated mutual information (MI) between residues from multiple sequence alignments of homologous proteins for identifying interacting pairs of residues in interacting proteins, and combined it with the discriminative random field (DRF), which is useful to extract some characteristic regions from an image in the field of image processing, and is a special type of conditional random fields (CRFs). In a similar way, in this paper, we make use of mutual information for predicting interactions between protein residues and RNA bases. Furthermore, we introduce labels of amino acids and bases as features of a simple two-dimensional CRF instead of DRF. To evaluate our method, we perform computational experiments for several interactions between Pfam domains and Rfam entries. The results suggest that the CRF model with MI and labels is more useful than the CRF model with only MI.

16:05-16:25 Coffee break

16:25-18:30 ISB Regular Session A1 (Conference Room Peony at Xi'an Jianguo Hotel)

Topic: Complex disease
Chair: Yan Zhang

16:25-16:50 *Using NMFAS to Identify Key Biological Pathways Associated With Human Diseases*

Hao Guo, Yunping Zhu, Dong Li, Fuchu He, Qijun Liu
State Key Laboratory of Proteomics, Beijing Proteome Research Centre, Beijing Institute of radiation Medicine, Beijing, P. R. China

Paper ID: 19

Abstract: Gene expression microarray enables us to measure the gene expression levels for thousands of genes at the same time. Here, we constructed the non-negative matrix factorization analysis strategy (NMFAS) to dig the underlying biological pathways related with various diseases by factorizing the pathway expression matrix, which was extracted from microarray matrix using pathway membership information, into the product of row and column vectors. We defined row vector as the pathway activity and column vector as the gene contribution weight. Via comparing the pathway activity of two different sample groups, we can identify significantly expressed pathways. We applied this strategy on two different cases: smoking and type 2 diabetes (DM2). We found 152 differentially expressed pathways by the comparison of pathway activity between smoker and never smoker, including pathways that have been validated in literature, such as "O-Glycans biosynthesis" and "Glutathione metabolism". We also found important genes related to smoking phenotype, such as NQO, HSPA1A, ALDH3A1. As for DM2 analysis, our results suggested 9 pathways were significantly expressed, including typical pathways like "Oxidative phosphorylation" and "mTOR signaling pathway", and found genes like CAPNS1, APP, COX7A1, COX7B, which might play important roles in the cellular regulations of DM2. In conclusion, Our strategy can be efficiently used to integrate gene expression profiles and biological pathway information to identify the key processes underlying human disease and can identify gene pathways missed by alternative approaches.

16:50-17:15 *Clinical Data Analysis Reveals Three Subytpes of Gastric Cancer*

Xinxin Wang, Duren Zhana, Chao Zhang, Lin Chen, **Yong Wang**
National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Paper ID: 75

Abstract: Gastric cancer is the fourth most common cancer and second leading cause of cancer-related death worldwide. Nowadays the accumulated large scale clinical data allows the clinicopathlogical review to identify the clinical factors, reveal their possible correlations, and mine the possible clinical patterns for gastric cancer. Here we analyze the clinical data of over 1500 gastric cancer patients histopathologically diagnosed and treated during 2006 to 2010. Specifically, we collect and preprocess the data by extracting 14 available clinical factors from three categories, i.e., the clinical background, immunohistochemistry data, and the caner's stage information. Then these factors are quantized and the significant factors and their correlations are calculated. Importantly, we define a distance between two patients by their clinical factors profile similarity and cluster all the patients into subgroups. We find that most of the patients fall into

three major classes and we define them as three subtypes of gastric cancer. Each subtype is analyzed and characterized by its own significant factors and correlations. Our analysis may provide important insights for gastric cancer classification and diagnose.

17:15-17:40 *in silico identification of novel cancer-related genes by comparative genomics of naked mole rat and rat*

Zhiyuan Yang, Yan Zhang, Luonan Chen
Center for systems biology, Soochow University, Suzhou, China
Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

Paper ID: 68

Abstract: The naked mole rat (NMR, Heterocephalus glaber) is a long-lived underground mammal, whose maximum lifespan can be up to 30 years and more than 7 times longer than house mouse. In addition, they are resistant to both spontaneous and experimentally induced tumorigenesis. These special biologic or behavioral characteristics make them most suitable for cancer and longevity research. The recent genome sequencing of NMR has provided the opportunity for the study of molecular mechanisms of such extreme traits. In this study, we carried out a comparative analysis of the complete set of NMR and rat genes. First, we identified all orthologous genes shared between these two animals. We further focused on the rat genes that were absent in NMR and used KEGG pathway database to identify the biological meaning of their proteins. The top three pathways include "Cytokine-cytokine receptor interaction", "Neuroactive ligand-receptor interaction" and "Pathways in cancer", which was consistent with the unique NMR traits. Interestingly, in the rat cancer pathway which contains 13 paths leading to evading apoptosis, 8 of them appeared to be interrupted in NMR. Finally, we found that 50% of genes lacked in "Pathways in cancer" and 40% of genes lacked in "MAPK signaling pathway" have been known to be related to a variety of cancers. Overall, this study provides insights into searching for new cancer-related genes and understanding the anti-cancer mechanism of NMR.

17:40-18:05 *Metabolite Biomarker Discovery For Metabolic Diseases By Flux Analysis*

Limin Li, **Hao Jiang**, Wai-Ki Ching, Vassilis S. Vassiliadis
Advanced Modeling and Applied Computing Laboratory Department of Mathematics The Univ ersity of Hong Kong, Pokfulam Road, Hong Kong

Paper ID: 1

Abstract: Metabolites can serve as biomarkers and their identification has significant importance in the study of biochemical reaction and signalling networks. Incorporating metabolic and gene expression data to reveal biochemical networks is a considerable challenge, which attracts a lot of attention in recent research. In this paper, we propose a promising approach to identify metabolic biomarkers through integrating available biomedical data and disease-specific gene expression data. A Linear Programming (LP) based method is then utilized to determine flux variability intervals, therefore enabling the analysis of significant metabolic reactions. A statistical approach is also presented to uncover these metabolites. The identified metabolites are then verified by comparing with the results in the existing literature. The proposed approach here can also be applied to the discovery of potential novel biomarkers.

18:05-18:30 *Identification of Oncogenic Genes for Colon Adenocarcinoma from Genomics Data*

Changhe Fu, Ling Jing, Su Deng, Guangxu Jin
College Of Science, China Agricultural University, Beijing, China

Paper ID: 64

Abstract: Identification of oncogenic genes from comprehensive genomics data with large sample size is of challenge. Here, we apply a well-established computational model, Bayesian factor and regression model (BFRM), to predict unknown colon cancer genes from colon adenocarcinoma genomic data. The BFRM takes advantages of its latent factors to characterize the underlying association between genes and the large number of colon cancer patients. Based on the known cancer genes in Online Mendelian Inheritance in Man (OMIM), we addressed three important latent factors focusing on characterization of heterogeneity of expression patterns related to specific oncogenic genes from the microarray data of 174 colon cancer patients. We found that the three latent factors can be employed to predict unknown colon cancer genes using the known oncogenic genes. These predicted unknown cancer genes were extensively validated by using the new somatic genes identified in the same patients from DNA sequencing data.

16:25-18:30 ISB Session B6 (Conference Room Pingpong at Xi'an Jianguo Hotel)

Topic: Dynamics Systems Biology

Chair: Katsuhisa Horimoto

16:25-16:50 *Multi-objective Optimization of Biological Systems Represented by S-system Models*

Gongxian Xu

Department of Mathematics, Bohai University, Jinzhou, China

Paper ID: 27

Abstract: This paper considers multi-objective optimization problems of biological systems. The biological system is represented by the S-system formalism. The advantage of this representation is that the steady-state equations are linear when the variables of the models are expressed in logarithmic coordinates. Profiting from this special property of S-system models, we transform the original nonlinear problem into a multiobjective linear programming. The obtained problem is then reformulated as a new multi-objective programming that has no equality or inequality constraints. The example of tryptophan biosynthesis is performed to the proposed framework and shown to the effectiveness of the approach. The simulation is also studied to give a performance comparison between the proposed and nonlinear approaches.

16:50-17:15 *System identification of the fermentation system of Thermoanaerobacter sp. X514*

Jing Yang, Xiaofang Ling, Lishan Yao, Hualiang Wei, Visakan Kadirkamanathan

The Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, P.R.China.

Paper ID: 42

Abstract: .Bioethanol production by means of anaerobic thermophilic microorganisms with pentose or hexose as the substrate are of paramount importance in sustainable fuel innovation. Manipulation of microorganisms and the associated experiment conditions by means of various ad-hoc technology is obviously the most straightforward way with the aim of maximizing bioethanol yield. However, methodology by means of mathematical modeling and analysis is often neglected among these routines. In this paper, typical input-output models are applied in the metabolic system analysis of Thermoanaerobacter sp. X514 under sole glucose substrate, sole xylose substrate and mixed glucose and xylose substrates conditions. Orthogonal Least Squares (OLS) approach is used for model parameter estimation. Model selection is proposed in order to testify the generality of the suggested model. System identification results illustrate that various forms of Nonlinear AutoRegressive with eXogenous input models (NARX) are applicable in delineating the system where different substrates (glucose or xylose) are utilized during the experiments. The proposed model structure infers that the yields of various products in X514 are mainly driven by the history information of the substrate consumption change. Moreover, the interaction between the main fermentation products of X514 is indirectly connected through the proposed models.

17:15-17:40 *Coupled Positive Feedback Loops Regulate the Biological Behavior*

Fei Shi, Peipei Zhou, Ruiqi Wang

Department of Mathematics, Institute of Systems Biology, Shanghai University, Shanghai 200444, China

Paper ID: 44

Abstract: Coupled positive feedback loops are frequentlyoccurring motifs in gene transcription regulatory networks and signaling pathways. So it's very important to investigate the function of coupled positive feedback loops. In this paper we establish mathematical models of coupled positive feedback loops. Through the bifurcation analysis, we prove that two coupled positive feedback loops can generate reversible and irreversible switch. And coupled positive feedback loops can strengthen bistable, enlarge signal and extend the signal reaction time. Coupled positive feedback loops play an important role in regulate biological behaviors.

17:40-18:05 *Alternating Weighted Least Squares Parameter Estimation for Biological S-Systems*

Li-Zhi Liu, **Fang-Xiang Wu**, Wen-Jun Zhang

Department of Mechanical Engineering and Division of Biomedical Engineering in the University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada.

Paper ID: 2

Abstract: The study of the global stability is essential for designing and controlling genetic regulatory networks. Most

existing results on this issue are based on linear matrix inequality (LMI) approach, which results in checking the existence of feasible solutions to high dimensional LMIs. In our previous study, we present several stability conditions for genetic regulatory networks with time-varying delays, based on M-matrix theory and the non-smooth Lyapunov function. In this paper, we design a smooth Lyapunov function and employ M-matrix theory to derive new stability conditions for genetic regulatory networks with time-varying delays. Theoretically, these conditions are less conservative than existing ones in some cases. For genetic regulatory networks with n genes and n proteins, these conditions become to check if an $n \times n$ matrix is an M-matrix, which is much easier than existing results. To illustrate the effectiveness of our theoretical results, two genetic regulatory networks are analyzed.

18:05-18:30 *Dynamics of Coexistence of Asexual and Sexual Reproduction in Adaptive Landscape*

Shuyun Jiao, Yanbo Wang, Ping Ao
Computational Biology Research Center, AIST, Tokyo, Japan
Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine of Ministry of Education, Shanghai Jiao Tong University, 200240, Shanghai, P.R.China

Paper ID: 77

Abstract: **Background:** The dynamics for species, especially rare species, with mixed type reproduction is instructive and meaningful for predicting the extinction of them. Though the extinction time is still a difficult problem for sexual or asexual population. Adaptive landscape introduced by Wright, a powerful concept in system biology can describe the evolution of organisms. To our knowledge, the dynamical of inhomogenous reproductive organisms have not investigated by a simple model globally. Methods: We describe how Wright-Fisher process maps to a mixed type population. We analytically construct adaptive landscape from the general diffusion equation. It shows that the construction is dynamical and the adaptive landscape is independent of the existence and normalization of the stationary distribution. **Results**:We first give a global model describing the evolution of a inhomogenous reproductive population by adaptive landscape. We visualize the dynamical behavior by adaptive landscape. Finite and infinite potential occur in the process. These results suggest a possible way to investigate the complex reproduction in a inhomogeneous reproductive population.

18:30-20:00 Tang dinner & Performance, Departure at 18:30 from lobby

August 20 (Monday) Technical Sessions

8:00-10:05 ISB Regular Session A2 (Conference Room Peony at Xi'an Jianguo Hotel)

Topic: Integrative Bioinformatics

Chair: Hyunju Lee

08:00-08:25 *Improving Prediction of Drug Therapy Outcome via Integration of Time Series Gene Expression and Protein Protein Interaction Network*

Liwei Qian, Haoran Zheng
School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, PR China.

Paper ID: 3

Abstract: Drug therapy to patients is often with partial success, and has been associated with a number of adverse reactions. Prediction of patients' response to therapy at the early stage of the treatment is crucial to avoiding those unnecessary adverse reactions. In this paper, a new approach that integrates time series gene expression and Protein Protein Interaction (PPI) network is presented to improve the prediction of patients' response to drug therapy. Experimental results showed that our method outperformed previous approaches. The method proposed here offers a huge potential for applications in pharmacogenomics and medicine.

08:25-08:50 *An Integrative Framework for Identifying Consistent MicroRNA Expression Signatures Associated with Clear Cell Renal Cell Carcinoma*

Jiajia Chen, Daqing Zhang, Wenyu Zhang, Yifei Tang, Lingchuan Guo, **Bairong Shen**
Center for Systems Biology, Soochow University, Suzhou, China
Paper ID: 9
Abstract: Clear cell renal cell carcinoma (ccRCC) is the most common and invasive renal-originated malignancy. Altered microRNA expression has been observed in many human cancers including ccRCC. Microarray is routinely used in labs worldwide for detecting cancer specific microRNA expression profiles, but no consistent conclusion could be drawn so far. The function of microRNAs in carcinogenesis of this tumor type is thereof largely unknown. In this study, we describe an integrative framework to improve the comparability of differentially expressed microRNAs (DE-miRNAs) from different experiments, and apply it to 4 publicly available microRNA expression datasets in ccRCC. The approach uses a novel statistic method for cancer outlier detection. The identified DE-miRNAs are then screened by POMA, an in-house developed predictor, for microRNAs with real regulatory activity in the disease. The proposed framework not only achieves high reproducibility across different datasets but also identifies a consistent set of 12 DE-miRNAs which could be putative biomarkers and therapeutic targets. The targets of DEmiRNAs in each dataset were then mapped to functional databases for enrichment analysis. Both novel and previously characterized microRNA-regulated molecular pathways are identified that are likely to contribute to the pathogenesis of ccRCC. Overlapping comparison suggests that independent ccRCC expression profiles are more consistent at pathway level than that at gene/microRNA level.

08:50-09:15 *Identifying novel glioma associated pathways based on integrated 'omics' data*
Yangfan Hu, Jinquan Li, Jiajia Chen, **Guang Hu**, Bairong Shen
Center for Systems Biology, Soochow University, Suzhou, China
Paper ID: 11
Abstract: Microarray represents a high throughput technology for analyzing expression profiles, and thus it has been widely applied in the study of pathogenesis of glioma. However, most of the analyses focused on detecting the differentially expressed genes in glioma. Although it is well accepted that the pathwayderived signatures is more reproducible than that at gene level, few meta-analysis of multiple microarray datasets at system level has been previously performed. In this article, we performed meta-analysis on different published glioma expression profiles and compared the overlapping of signature at gene and pathway level. Pathway enrichment analysis result of GeneGO database and Gene Set Enrichment Analysis (GSEA) showed that 100% and 64% of the similarity was higher than that of genes respectively. Moreover, we integrated other omics data on glioma, such as MicroRNA expression profiles and Chip-Seq data, for further verification. The results showed that the significant signatures of different data sets are more similar at pathway level than at gene level. 12 pathways found by GeneGO database were shared by four stages among several datasets. 5 of these pathways, such as Regulation of epithelialto- mesenchymal transition (EMT), TGF-beta-dependent induction of EMT via SMADs, were putative novel pathways on glioma and need further experimental verification.

09:15-09:40 *A Novel Pipeline for Motif Discovery, Pruning and Validation in Promoter Sequences of Human Tissue Specific Genes*
Xiu-Jun Gong, Hua Yu, Fei-Fei Zhao
School of Computer Science and Technology, Tianjin University, Tianjin, China
Paper ID: 43
Abstract: Identification and analysis of tissue-specific (TS) genes and their regulatory activities play an important role in the understanding of mechanisms of organisms, disease diagnosis and drug design. In this paper, we designed a pipeline for the discovery of promoter motifs for tissue-specific genes. The pipeline consists of three phases: motif searching, motif merging and motif validation. The motif searching phase integrated three algorithms: MEME, AlignACE and Gibbs Sampling. In the second phase, we proposed a motif merging method, which is based on Bayesian probabilistic principles, to reduce redundancies of motifs from the first phase. Lastly, the motif validation phase verified the statistical significance of discovered motifs using a Bayesian Hypothesis Test approach. We performed the analysis on the sequences of promoter regions (-449bp-1000bp) of 4,552 human tissue-specific genes across 82 tissues and 924 housekeeping genes. The distributions of motifs in different promoter regions show that most motifs prefer to be in the proximal region (+500~50bp, -50bp~-500bp) of promoters.

09:40-10:05 *Predicting protein complexes via the integration of multiple biological information*
Xiwei Tang, **Jianxin Wang**, Yi Pan
AMAC Laboratory, Department of Mathematics, The University of Hong Kong, Hong Kong, China
Paper ID: 45

Abstract: Protein complexes are a cornerstone of many biological processes and together they form various types of molecular machinery that perform a vast array of biological functions. An increase in the amount of protein-protein interaction (PPI) data enables a number of computational methods for predicting protein complexes. There are a mass of algorithms detecting complexes only consider the PPI data. However, the PPI data from high-throughout techniques is flooded with false interactions. In fact, the insufficiency of the PPI data significantly lowers the accuracy of these methods. In the current work, we develop a novel method named CMBI to discover protein complexes via the integration of multiple biological resources including gene expression profiles, essential protein information and PPI data. First, CMBI defines the functional similarity of each pair of interacting proteins based on the edge-clustering coefficient (ECC) from the PPI network and the Pearson correlation coefficient (PCC) from the gene expression data. Second, CMBI selects essential proteins as seeds to build the protein complex cores. During the growth process, the seeds' essential protein neighbors and the neighbors whose functional similarity (FS) with the seeds are more than the threshold T will be added to the complex cores. After the complex cores are constructed, CMBI begins to generate protein complexes by attaching their direct neighbors with FS > T to the cores. In addition to the essential proteins, CMBI also uses other proteins as seeds to expand protein complexes. To check the performance of CMBI, we compare the complexes discovered by CMBI with the ones found by other techniques by matching the predicted complexes against the reference complexes. We use subsequently GO::TermFinder to analyze the complexes predicted by various methods. Finally, the effect of parameter T is investigated. The results from GO functional enrichment and matching analyses show that CMBI performs significantly better than the state-of-the-art methods. It means that it's successful for us to integrate multiple biological information to identify protein complexes in the PPI network.

08:00-10:05 ISB Regular Session B7 (Conference Room Pingpong at Xi'an Jianguo Hotel)

Topic: Dynamics Systems Biology
Chair: Kazuhiko Fukui

08:00-08:25 *Comparing two models based on the transcriptional regulation by KaiC of cyanobacteria rhythm*

Ying Li, Hui Wu, Jinhuo Luo

College of Information Technology, Shanghai Ocean University, Shanghai 201306

Paper ID: 53

Abstract: Circadian clocks are self-sustained biological oscillators that can be entrained by environmental cues. Cyanobacteria are the simplest organisms known to exhibit circadian rhythms, which is the fundamental process of homeostasis adapting to daily environmental changes. The cyanobacterial clock gene products, KaiA, KaiB, and KaiC interact with each other, and regulate KaiC phosphorylation and kaiBC expression in a circadian fashion. The total phosphorylation level of KaiC oscillates with a circadian period. In this paper, based on two possible transcriptional regulations, we examined numerically two models, the Transcriptional Activation Model and the Transcriptional Repression Model to generate circadian oscillation of kai genes. These two models both reproduce experimental observed sustained circadian oscillations in constant dark(DD) and constant light(LL). Comparing phase shifts between DD and LL in these two model, the Transcriptional Activation Model is consistent with the experimental observations, suggesting that the Transcriptional Activation Model may reflect the essence of the actual mechanism of kai oscillator in cyanobacteria.

08:25-08:50 *New Global Stability Conditions for Genetic Regulatory Networks with Time-Varying Delays*

Li-Ping Tian, Zhong-Ke Shi, Fang-Xiang Wu

School of Information, Beijing Wuzi University, Beijing, P. R. China.

Paper ID: 48

Abstract: The study of the global stability is essential for designing and controlling genetic regulatory networks. Most existing results on this issue are based on linear matrix inequality (LMI) approach, which results in checking the existence of feasible solutions to high dimensional LMIs. In our previous study, we present several stability conditions for genetic regulatory networks with time-varying delays, based on M-matrix theory and the non-smooth Lyapunov function. In this paper, we design a smooth Lyapunov function and employ M-matrix theory to derive new stability conditions for genetic regulatory networks with time-varying delays. Theoretically, these conditions are less conservative than existing ones in some cases. For genetic regulatory networks with n genes and n proteins, these conditions become to check if an $n \times n$ matrix is an M-matrix, which is much easier than existing results. To illustrate the effectiveness of our theoretical results, two genetic regulatory networks are analyzed.

08:50-09:15 *Module network rewiring in response to therapy*

Tao Zeng, Ruozhao Wang, Luonan Chen
Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre, for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

Paper ID: 50

Abstract: Response to stress is an important biological mechanism to react to environment variations. Different from distinguishing stresses like heat shock, ER stress, and oxidative stress, the study of response to an artificial signal like drug in therapy would be an alternative and also attractive way to understand the cellular response mechanism, which also benefits clinical application. Although differentially expressed genes are usually thought to be therapy responsive genes in many previous researches, more and more attention is diverted from single genes to functions or pathways, in particular for cancer therapy analysis. Thus, comparing with purely molecule (e.g., gene) rewiring, understanding functional reorganization or module rewiring would be more important for systematically studying therapy response or other dynamic biological processes. Therefore, in this paper we propose a model of module network rewiring to characterize functional reorganization, in contrast to gene network rewiring. Specifically, we develop a new framework named as module network rewiring analysis (MNRA) to investigate relevant network modules and their re-connections during an antiviral therapy. In MNRA, we aim to study module dynamics from the network viewpoint, by defining a module network with a module as a node and a path connecting two modules as an edge, which is a network for the molecular interaction system on a higher level. By MNRA experiments on expression data of patients with Hepatitis C virus infection (HCV) receiving Interferon therapy, we found that (1) the consistent module (a set of genes) separates two new subtypes of patients which were not discovered by differentially expressed genes; (2) the patient-group specific module network rewiring reveals necessary functional connections bridged by biological paths; (3) the hierarchical structures of temporal module network rewiring show that they can be taken as spatial-temporal markers to diagnose whether a patient has therapy response or not. Thus, MNRA indeed can provide biologically systematic clues for potential pharmacogenomic applications and has ability to characterize complex dynamic processes for many biological systems.

09:15-09:40 *Dynamic miRNA-TF-mRNA circuits in mouse lung development*

Xinghuo Ye, **Juan Liu**, Fang-Xiang Wu
School of Computer, Wuhan University, Wuhan 430072, China

Paper ID: 63

Abstract: Genes, transcription factors (TF), microRNAs (miRNA) are well-known to have important regulating roles in dynamic biological processes. In the last years, many studies have been devoted to the elucidation of transcriptional or posttranscriptional regulating activities of TFs or miRNAs, respectively. However, very limited attempts have been made to consider the dynamic characteristics of miRNA-TF-mRNA circuits, which are the biological network motifs considering miRNAs, TFs and genes as a whole in the complicated biological procedures like mouse lung development. Here we propose to mine miRNA-TF-mRNA circuits related to the mouse lung development by integrating TF-mRNA, miRNA-mRNA, TFmiRNA, and time-course expression data, and to further analyze the variations of these circuits in different stages of the lung development. To our best knowledge, this is the first time to take transcriptional and post-transcriptional information together to describe the mouse lung development. Our preliminary results show that miRNA-TF-mRNA circuits vary in different stages of the lung development and play different roles.

09:40-10:05 *Analysis of A HBV Infection Models with ALT*

Yongmei Su, Yerong Wen, Lequan Min
School of Mathematics and physics,University of Science and Technology Beijing Beijing,100083, PR China

Paper ID: 28

Abstract: Mathematical models have been used to understand the factors that govern infectious disease progression in viral infections. Many hepatitis B virus (HBV) models were set up based on the basic virus infection model (BVIM) introduced by Zeuzem et al. and Nowak et al. But some references have pointed out that the basic infection reproductive number of the BVIM is biologically questionable and given the modified models. And so far, no immune model with alanine aminotransferase (ALT) was given based on the modified models. In this paper one immune models with ALT based on the modified model is discussed. The stability analysis and simulation of the model is also given based on clinical data of ALT and HBV DNA.

10:05-10:25 Coffee break

10:25-12:30 ISB Regular Session A3 (Conference Room Peony at Xi'an Jianguo Hotel)

Topic: Bioinformatics
Chair: Shwu-Rong Shieh

10:25-10:50 *Human Encoded miRNAs that Regulate the Inflenenza Virus Genome*

Hao Zhang, Xin Li, Yuaning Liu, Minggang Hu, Zhi Li, Dong Xu
Symbol Computation and Knowledge Engineering of Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun , China

Paper ID: 5

Abstract: Motivation:MiRNAs can downregulate gene expression by mRNA cleavage or translational repression. Discovering human encoded miRNAs that regulate the influenza virus genome is important for molecular targets for drug development, and it also plays positive role in influenza control and prevention. Methods: We propose a new method based on scoring to discover human encoded miRNAs that regulate the influenza virus genome. The scoring based on the same complementary sites, the secondary structure of the complementary sites and the bind sites of all sequences respectively. Among them, taking the secondary structure as a vital factor is a new attempt. Results: Has-miR-489, has-miR-325, has-miR-876-3p and hasmiR-2117 are targeted HA, PB2, MP and NS of influenza A, respectively.

10:50-11:15 *A Novel Information Contents Based Similarity Metric for Comparing TFBS Motifs*

Shaoqiang Zhang, Lifen Jiang, Chuanbin Du, Zhengchang Su
College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China

Paper ID: 7

Abstract: Identifying binding sites recognized by transcription factors (TFs) is one of major challenges to decipher complex genetic regulatory networks encoded in a genome. A set of binding sites recognized by the same TF, called a motif, can be accurately represented by a position frequency matrix (PFM) or a position-specific scoring matrix (PSSM). Very often, we need to compare motifs when searching for similar motifs in a motif database for a query motif, or clustering motifs possibly recognized by the same TF. In this paper, we have designed a novel metric, called SPIC (Similarity between Positions with Information Contents), for quantifying the similarity between two motifs using their PFMs, PSSMs, and column information contents, and demonstrated that this metric outperforms the other state-of-the-art methods for clustering motifs of the same TF and differentiating motifs of different TFs.

11:15-11:40 *A fixed-point blind source extraction algorithm and its application to ECG data analysis*

Hongjuan Zhang, **Zikai Wu**, Shuxue Ding, Luonan Chen
Department of Mathematics, Shanghai University, Shanghai,200444,P R China

Paper ID: 18

Abstract: Generalized autocorrelations and complexity pursuit are two recently developed methods for extracting interesting component from time series. They are the extensions of projection pursuit to time series data. In this paper, a fixedpoint blind source extraction (BSE) algorithm for generalized autocorrelations and complexity pursuit of the desired signals is presented. The fixed-point algorithm inherits the advantages of the well-known FastICA algorithm of ICA, which is very simple, converges fast, and does not need to choose any learning step sizes. Numerical experiments on electrocardiogram (ECG) data indicate its better performance.

11:40-12:05 *Anti-clustering of circadian gene expression in mouse liver genome*

Bin Kang, **Yuan-yuan Li**, Yi-xue Li
Shanghai Center for Bioinformation Technology, Shanghai 200235, China

Paper ID: 66

Abstract: Circadian regulatory system is an evolutionarily ancient biological system. Its prevalence in life kingdoms suggests it has fundamental role in life processes. Although genomic scale of circadian gene expression has been found in

various species from cyanobacteria to mammalians, transcriptional patterns and mechanisms of global circadian gene regulation have not yet been revealed. Using high resolution temporal profiling of mouse circadian gene expression, we show that contrary with previously demonstrated clustering tendency of functionally related genes in mammalian genomes, circadian regulated genes display anti-clustering propensity in mouse liver. This unique property does not conform to the notion of domain-wide coordinated gene regulation dictated by acetyl modifications, which is recently identified as a hallmark of circadian regulation. These results suggest that global circadian regulation in mouse liver might involve other structural chromosome interactions irrelevant with clustering regulation.

11:40-12:05 P*redicting biological dynamics based on short high throughput time-course data*

Huanfei Ma, Kazuyuki Aihara, and Luonan Chen

Institute of Industrial Science, The University of Tokyo, Japan

Abstract: In this paper, the prediction of biological dynamics for high dimensional data is considered. Particularly, the prediction problem for the short high throughput time-course data will be investigated, which is believed to be a hard problem due to the conventional wisdom. A method based on the attractor embedding theory is proposed and the algorithm for a specific situation is designed to calculate the embedding map. With the embedding map, the predictions will be made for the time-course data. Both theoretical analysis and real data validations are investigated.

10:25-12:30 ISB Regular Session B8 (Conference Room Pingpong at Xi'an Jianguo Hotel)

Topic: Bioinformatics

Chair: Ning Kang

10:25-10:50 *A New Method to Identify Repositioned Drugs for Prostate Cancer*

Zikai Wu, Yong Wang, Luonan Chen

University of shanghai for Science and Technology, Shanghai, 200093 China

Paper ID: 67

Abstract: With the merits of faster development time and reduced risk, identifying new indications for marketed drugs draws more and more attention. In particular, repositioning drugs with known indications has become an hot topic in the area of computational systems biology. However, one of the common shortcomings for most of the previous methods is the ignorance of side effect, i.e., drug through primary targets and off targets might induce both desired and unintended effects respectively, which could not appropriately evaluated in most of existing methods. In this paper with a new measure considering both efficacy and side effect, we developed a new method for identifying the repositioned drugs against prostate cancer by evaluating the mutual relations of the gene expression levels between prostate cancer samples and those induced by bioactive compounds. In this measure, the overlap between gene sets that were oppositely regulated in disease state and drug treatment state was quantified by jaccard index as drug's efficacy while the overlap between essential genes and positively correlated genes (or regulated just after drug treatment) was quantified by jaccard index as drug's side effect, which were balanced with a parameter \lambda. The preliminary results on repositioning drugs for prostate cancer verify the effectiveness and efficiency of the new method.

10:50-11:15 *Application of Granger Causality to Gene Regulatory Network Discovery*

Gary Hak Fui Tam, Chunqi Chang, Yeung Sam Hung

Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Paper ID: 57

Abstract: Granger causality (GC) has been applied to gene regulatory network discovery using DNA microarray time-series data. Since the number of genes is much larger than the data length, a full model cannot be applied in a straightforward manner, hence GC is often applied to genes pairwisely. In this paper, firstly we investigate with synthetic data and point out how spurious causalities (false discoveries) may emerge in pairwise GC detection. In addition, spurious causalities may also arise if the order of the vector autoregressive model is not high enough. Therefore, besides using a suitable model order, we recommend a full model over pairwise GC. This is possible if pairwise GC is first used to identify a network of interactions among only a few genes, and then all these interactions are validated with a full model again. If a full model is not possible, we recommend using model validation techniques to remove spurious discoveries. Secondly, we apply pairwise GC with model validation to a real dataset (HeLa). To estimate the model order, the Akaike information criterion is found to be more suitable than the Bayesian information criterion. Degree distribution and network hubs are obtained and compared with previous publications. The hubs tend to act as sources of interactions rather than receivers of interactions.

11:15-11:40 *A Novel Feature Selection Method Based on CFS in Cancer Recognition*

Xinguo Lu, Xianghua Peng, Ping Liu, **Yong Deng**, Bingtao Feng, Bo Liao
School of Information Science and Engineering, Hunan University, Changsha, 410082, China
Paper ID: 56
Abstract: In recent years, the gene expression profiles are used for cancer recognition. But the researchers are disturbed by their large variables and small observes. In this paper, a novel feature selection method based on correlation-based feature selection(CFS) was proposed. Firstly, the measures of variable to variable and variable to observe were calculated respectively. Then we utilized heuristic search method to search the space of variable for selecting informative gene subset and the subset weight was computed using these measures. Through regression we obtained a subset of distinguished genes. Finally, the stratified sampling strategy was presented to obtain the most informative genes. And classification performance was tested to evaluate the proposed method. Ten-fold cross-validation experiment was performed in three datasets including leukemia, colon cancer and prostate tumor. The experimental results show that the proposed method can obtain the distinguished gene subset and different classifier can acquire better classification performance with this subset.

11:40-12:05 *Sparse Kernel Logistic Regression for β-turns Prediction*
Murtada Khalafallah Elbashir, **Jianxin Wang**, FangXiang Wu, Min Li
School of Information Science and Engineering, Central South University, Changsha, 410083, P.R. China.
Paper ID: 59
Abstract: A beta-turn is a secondary protein structure type that plays a significant role in protein folding, stability, and molecular recognition. On average 25% of amino acids in protein structures are located in beta -turns. Development of accurate and efficient method for beta -turns prediction is very important. Most of the current successful beta -turns prediction methods use support vector machines (SVMs) or Neural Networks (NNs), however a method that can yield probabilistic outcome, and has a well-defined extension to the multi-class case will be more valuable in beta –turns prediction. Although kernel logistic regression (KLR) is a powerful classification technique that has been applied successfully in many classification problems, however it is often not found in beta -turns classification, mainly because it is computationally expensive. In this paper we used KLR to obtain sparse beta –turns prediction in short evolution time after speeding it using Nystrom approximation method. Secondary structure information and position specific scoring matrices (PSSMs) are utilized as input features. We achieved Qtotal of 80.4% and MCC of 50% on BT426 dataset. These results show that KLR method with the right algorithm can yield performance equivalent or even better than NNs and SVMs in beta-turns prediction. In addition KLR yields probabilistic outcome and has a well-defined extension to multi-class case.

12:30-13:20 Lunch break (Rongyuan Restaurant at Xi’an Jianguo Hotel)

Social Program: Half-day tour

13:20-18:30 Half day excursion in The Terra Cotta Warrior & House Museum. Departure at Xi’an Jianguo Hotel Lobby

18:30-22:00 Banquet at Xi'an Grant Restaurant

*The above program subjects to revision based on further information and Ad Hoc presentation requests.

Prof. Luhua Lai

Prof. Luhua Lai is the Vice Dean and Changjiang Professor, College of Chemistry and Molecular Engineering in Peking University. She is also the Director of The State Key Laboratory for Structural Chemistry of Unstable and Stable Species. She got her B.Sc.in 1984, from Peking University, M.Sc. in 1987 from Peking University, and Ph.D.in 1989 from Peking University. During 1998-1999, she was the Berkeley Scholar in the University of California at Berkeley. Professor Lai's group focuses on solving biological problems by using physical chemistry, especially structural chemistry methods. Related to: chemical biology, biophysical chemistry, structural chemistry, structural biology, bioinformatics, computational chemistry, computational and theoretical biology.

http://mdl.ipc.pku.edu.cn/

Prof. Masahiro Okamoto

Prof. Masahiro Okamoto is the professor in Department of Bioinformatics, Graduate School of Systems Life Sciences in Kyushu University. He got his Ph.D. in Biochemistry from Department of Systems Life Sciences in the Graduate School of Systems Life Sciences. His research interests include Mathematical and Computational Strategies for Systems and Synthetic Biology and Analysis of bio-functional network and control by using information technology and practical implementation in biotechnological or engineering fields.

http://www.brs.kyushu-u.ac.jp/~okahon/

Prof. Raul Rabadan

Prof. Raul Rabadan is an Assistant Professor in the Department of Biomedical Informatics and in the Center for Computational Biology and Bioinformatics, at the Columbia University College of Physicians and Surgeons. Dr. Rabadan has been the Martin A. and Helen Chooljian Member at The Simons Center for Systems Biology at the Institute for Advanced Study in Princeton, New Jersey. From 2001 to 2003 he was a fellow at the Theoretical Physics Division at CERN, the European Organization for Nuclear Research, in Geneva, Switzerland. In 2003 he joined the Physics Group of the School of Natural Sciences at the Institute for Advanced Study. Dr. Rabadan's current interest focuses on patterns of evolution in biological systems—in particular, RNA viruses and cancer.

http://rabadan.c2b2.columbia.edu/

Prof. Yang Zhang

Prof. Yang Zhang, Ph.D. is an Associate Professor in Biological Chemistry and Computational Medicine and Bioinformatics in University of Michigan. He obtained his Ph.D., in Central China Normal University and was a Postdoc in the University at Buffalo. Prof. Zhang's interest is to understand the fundamental relations between protein sequence, structure and function. The major focus of his lab is to develop new bioinformatics algorithms to predict 3-dimensional protein structure from the amino acid sequence and deduce the biological function of proteins by comparing the predicted structures with the function databases. A number of computational methods developed by the Zhang lab have been demonstrated in the CASP experiments to be the world's best for protein structure prediction and function annotation. The lab is currently working on extending the developed protein modeling algorithms for protein design and structure-based drug discovery. They are especially interested in modeling G protein-coupled receptors and the interactions with the associated ligands with the purpose of developing new drugs to regulate these interactions

http://zhanglab.ccmb.med.umich.edu/

2012 IEEE
6th International Conference on Systems Biology (ISB)

Xi'an, China, August 18–20, 2012

Edited by

Luonan Chen
Xiang-Sun Zhang
Ling-Yun Wu
Yong Wang







2012 IEEE International Conference on Systems Biology (ISB)

IEEE Catalog Number CFP12ISB-ART
ISBN 978-1-4673-4398-5

Additional copies of this publication are available from

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
+1 845 758 0400
+1 845 758 2633 (FAX)
Email: curran@proceedings.com

ABOUT ISB 2012

THEME AND SCOPE

The IEEE 6th International Conference on Systems Biology (ISB 2012), organized by Chinese Academy of Sciences and Xidian University will be held in Xi'an, China, August 18–20, 2012. The conference is sponsored by National Natural Science Foundation of China (NSFC), Academy of Mathematics and Systems Sciences of CAS (AMSS), Shanghai Institutes for Biological Sciences of CAS (SIBS), Xidian University, K. C. Wong Education Foundation, Computational Systems Biology Society of ORSC, Systems Biology Technical Committee of IEEE SMC Society, and also sponsored by IET.

Systems Biology and Bioinformatics have become intensive research topics in the recent past decade and attracted great many leading scientists working in Biology, Physics, Mathematics and Computer Science. Optimization, Statistics, and many other mathematical methods have been widely used in the field. Following the successful OSB 2007–2009 and ISB 2010–2011, the purpose of ISB 2012 is to extend the international forum for scientists, researchers, educators, and practitioners to exchange ideas and approaches, to present research findings and state-of-the-art solutions in this interdisciplinary field, including mathematical methods and its applications in biosciences and researches on various aspects of Systems Biology, such as integration of genome-wide microarray, proteomic, and metabolomic data, inference and comparison of biological networks, and model testing through design of experiments.

The purpose of ISB 2012 is to provide an international forum for scientists, researchers to exchange ideas and approaches, including theoretical methodology development and its applications in biosciences and researches on various aspects of Computational Systems Biology. Themes of the ISB 2012 will be interdisciplinary by its nature and focus on bridging opportunities between mathematical methods and Systems Biology studies. We are particularly interested in submissions that report on theoretical, experimental and applied research motivated by systems biology problems. Typical, but not exclusive, topics of interest are:

- Gene Regulatory Networks
- Protein Interaction Networks
- Metabolic Networks
- Signaling Networks
- Comparative Genomics
- Functional Genomics
- Metagenomics
- Genome-Wide Association Study
- Promoter Analysis and Discovery
- Biomarker Identification and Drug Discovery
- Evolution and Phylogenetics
- Non-coding RNAs
- Proteomics
- Protein Structures and Functions
- Microbial Community Analysis
- Qualitative Analysis of Biological Systems
- Quantitative Models of Cellular and Multi-Cellular Systems
- Designing and Modeling Synthetic Biological Systems
- Nonlinear Dynamics and Analysis of Biological Systems
- Designing Synthetic Biological Circuits
- High Performance Computing for Biological Data Analysis
- Data Mining and Machine Learning for Biological Data
- Information Theory and Statistical Analysis
- Systems Biology of Cancer and Metastasis
- Brain Systems Biology
- Systems Neuro-Informatics
- Systems Biology of Development

COMMITTEES

- General Chairs
  - Luonan Chen (Chinese Academy of Sciences, China)
  - Xiang-Sun Zhang (Chinese Academy of Sciences, China)
- Program Chairs
  - Graziano Chesi (University of Hong Kong, Hong Kong)
  - Lin Gao (Xidian University, China)
- Publication Chairs
  - Ling-Yun Wu (Chinese Academy of Sciences, China)
  - Katsuhisa Horimoto (National Institute of Advanced Industrial Science and Technology, Japan)
- Highlights Track Chairs
  - Shihua Zhang (Chinese Academy of Sciences, China)
  - Junwen John Wang (The University of Hong Kong, Hong Kong)
- Organizing Committee Chairs
  - Degang Liu (Chinese Academy of Sciences, China)
  - Yong Wang (Chinese Academy of Sciences, China)
  - Xing-Ming Zhao (Shanghai University, China)

PROCEEDING PAPERS AND CONTRIBUTING AUTHORS

Sixty full papers in this volume cover wide range of computational systems biology. Authors of these papers come from China mainland, Hong Kong, Taiwan, Australia, Canada, Japan, Korea, Poland, Singapore, Thailand, United Kingdom, United States. Many active researchers in various areas contributed their overview and introduction in their fields besides specific deep research achievements.

Contents

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

978-1-4673-4398-5/12/$31.00 ©2012 IEEE