The 10th International Conference on Systems Biology (ISB 2016)







August 19-22, 2016 Weihai, China

ISB 2016 Schedule

August 19 Friday	13:00-18:00	Registration (Hotel Lobby)		
	18:00-20:00	Reception		
	20:00-21:30	Board member meeting of ORSC-CSB (Room WT)		
August 20 Saturday	08:40-08:50	Opening Session (Room AC)		
	08:50-10:30	Plenary Session P1 (Room AC)		
	10:30-10:50	Coffee break		
	10:50-12:30	Highlight Session A1 (Room AC)	Session B1 (Room HY)	
		Systems Biology	Drug and Disease	
		Paper IDs: 1, 11, 43, 61	Paper IDs: 3, 19, 25, 28	
		Chair: Tianshou Zhou	Chair: Fengfeng Zhou	
	12:30-13:30	Lunch		
	14:00-15:40	Highlight Session A2 (Room AC)	Session B2 (Room HY)	
		Biomarker and Drug Target	Network Biology	
		Paper IDs: 2, 4, 48, 58	Paper IDs: 15, 38, 40, 45	
		Chair: Bairong Shen	Chair: Yong Wang	
	15:40-16:20	Coffee break & Poster Session		
	16:20-18:00	Highlight Session A3 (Room AC)	Session B3 (Room HY)	
		Bioinformatics	Proteomics	
		Paper IDs: 36, 37, 60	Paper IDs: 14, 44, 46, 47	
		Chair: Yu Xue	Chair: Zhiping Liu	
	18:00-20:00	Banquet		

ISB 2016 Schedule

	06:30-08:30	The beach football match dedicated to Tenth Anniversary of ISB		
August 21 Sunday	08:50-10:30	Highlight Session A4 (Room AC)	Session B4 (Room HY)	
		Bioinformatics	Biomarker	
		Paper IDs: 54, 56, 59	Paper IDs: 26, 49, 53, 57	
		Chair: Xing-Ming Zhao	Chair: Shihua Zhang	
	10:30-10:50	Coffee break		
	10:50-12:30	Session A5 (Room AC)	Session B5 (Room HY)	
		Bioinformatics	Systems Biology	
		Paper IDs: 16, 20, 30, 31	Paper IDs: 27, 29, 39, 50	
		Chair: Minghua Deng	Chair: Kang Ning	
	12:30-13:30	Lunch		
	14:00-15:40	Plenary Session P2 (Room AC)		
	15:40-16:20	Coffee break & Poster Session		
	16:20-18:00	Session A6 (Room AC)	Session B6 (Room HY)	
		Bioinformatics	Systems Biology	
		Paper IDs: 7, 24, 41, 42	Paper IDs: 10, 21, 23, 34	
		Chair: Lei Li	Chair: Jingzhi Lei	
	18:00-20:00	Dinner		
August 22 Monday	8:00-13:00	Half day excursion in Liugong Island. Departure at 8:00 from lobby.		

Room AC: Academic Conference Room at 2nd floor (二楼学术报告厅)

Room HY: Haina-Yunsong Conference Room at 1st floor (一楼海纳云松会议厅)

Room WT: Wentao Meeting Room at 2nd floor (二楼文涛会议室)

ISB 2016 Program

August 19-22, Weihai, Shandong, China

August 19 (Friday) Registration

- **13:00-18:00 Registration**, Participants arrival in Weihai, check in hotel, and registration package pick up (Hotel Lobby at 2nd floor 二楼宾馆大堂).
- 18:00-20:00 Reception
- 20:00-21:30 Board member Meeting for Computational Systems Biology Society of ORSC (Wentao Meeting Room at 2nd floor 二楼文涛会议 室)

August 20 (Saturday) Technical sessions

08:30-11:00 Registration for late arrivals (Academic Conference Room at 2nd floor 二楼学术报告厅)

08:40-08:50 Opening Session (Academic Conference Room at 2nd floor 二楼学 术报告厅)

Chair: Luonan Chen

08:50-10:30 Plenary Session P1 (Academic Conference Room at 2nd floor 二 楼学术报告厅) Chair: Luonan Chen 8:50-09:40 Structure-Function Mapping of 3D Human Genome

Xianghong Jasmine Zhou
 University of California Los Angeles, USA

 09:40-10:30 Transcript-Based Differential Expression Analysis for Population RNA-Seq
 Data
 Tao Jiang
 University of California Riverside, USA

10:30-10:50 Coffee break

10:50-12:30 Highlight Session A1 (Academic Conference Room at 2nd floor 二 楼学术报告厅)

Topic: Systems Biology Chair: Tianshou Zhou

10:50-11:15 Molecular dynamics simulation reveals how phosphorylation of tyrosine 26 of phosphoglycerate mutase 1 upregulates glycolysis and promotes tumour growth Yan Wang and Guanyu Wang South University of Science and Technology of China PaperID: 1
11:15-11:40 Mathematical model reveals competitive roles of autophagy in inhibiting progression and promoting growth in pancreatic cancers Jinzhi Lei Tsinghua University PaperID: 11
11:40-12:05 Modeling of the dynamics of hematopoiesis system in cyclical thrombocytopenia Jinzhi Lei and Changjing Zhuge Tsinghua University

PaperID: 43 12:05-12:30 Robustness and controllability of complex networks Xueming Liu and Linqiang Pan Huazhong University of Science and Technology PaperID: 61

10:50-12:30 Session B1 (Haina-Yunsong Conference Room at 1st floor 一楼海 纳云松会议厅)

Topic: Drug and Disease Chair: Fengfeng Zhou

10:50-11:15 An Improved Feature-Based Approach to Predict Effective Drug *Combinations* Qian Xu, Yi Xiong, Hao Dai, Kotni Meena, Qin Xu, Hongyu Ou and Dongqing Wei Shanghai Jiaotong University PaperID: 3 **11:15-11:40** *Drug target prediction by multi-view low rank embedding* Menglan Cai and Limin Li Xi'an Jiaotong University PaperID: 19 **11:40-12:05** *Network-based method to understanding the therapeutic effect of* ShexiangBaoxin Pill on cardiovascular diseases Jing Zhao Logistical Engineering University PaperID: 25 12:05-12:30 Dynamics evolutions of HIV infection on treatment with combinatorial drugs Xiaojing Lu, Yanwei Liu and Ruiqi Wang Shanghai University PaperID: 28

12:30-13:30 Lunch

14:00-15:40 Highlight Session A2 (Academic Conference Room at 2nd floor 二 楼学术报告厅) Topic: Biomarker and Drug Target

Chair: Bairong Shen

14:00-14:25 Identifying network-based biomarkers of complex diseases from highthroughput data Zhi-Ping Liu Shandong University PaperID: 2
14:25-14:50 The network biomarker for complex diseases Tao Zeng Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 4

14:50-15:15 *Computational probing protein–protein interactions targeting small molecules*

Yongcui Wang, Shilong Chen, Naiyang Deng and Yong Wang Northwest Institute of Plateau Biology, Chinese Academy of Sciences PaperID: 48

15:15-15:40 Integrative analysis for identifying joint modular patterns of geneexpression and drug-response data

Jinyu Chen and Shihua Zhang

Academy of Mathematics and Systems Science, Chinese Academy of Sciences PaperID: 58

14:00-15:40 Session B2 (Haina-Yunsong Conference Room at 1st floor 一楼海 纳云松会议厅)

Topic: Network Biology Chair: Yong Wang

14:00-14:25 Accelerated Parallel Algorithm for Gene Network Reverse Engineering Jing He, Zhou Zhou, Michael Reed and Andrea Califano Columbia University, USA Dependent 15

PaperID: 15

14:25-14:50 Network clustering analysis using Mixture Exponential-family Random Graph Models and its application in Genetic Interaction

Yishu Wang, Huaying Fang, Dejie Yang, Minghua Deng and Hongyu Zhao Qingdao University

PaperID: 38

14:50-15:15 Meta-analysis for Feature Selection Based on Gene Co-expression Network Xue Jiang, Han Zhang and Xiongwen Quan Nankai University

PaperID: 40

15:15-15:40 Bayesian network model for identification of pathways by integrating protein interaction with genetic interaction data

Changhe Fu, Guangxu Jin, Su Deng and Zu-Guo Yu Shenyang Normal University PaperID: 45

15:40-16:20 Coffee break & Poster Session

16:20-18:00 Highlight Session A3 (Academic Conference Room at 2nd floor 二 楼学术报告厅)

Topic: Bioinformatics Chair: Yu Xue 16:20-16:45 BDB: biopanning data bank Bifang He, Guoshi Chai, Yaocong Duan and Jian Huang University of Electronic Science and Technology of China PaperID: 36
16:45-17:10 Data science – part of integrated agriculture solution Le Lv and Jingdong Liu Monsanto Company PaperID: 37
17:10-17:35 Minimize the number of features for a biomedical data mining problem Fengfeng Zhou Jilin University PaperID: 60

16:20-18:00 Session B3 (Haina-Yunsong Conference Room at 1st floor 一楼海 纳云松会议厅)

Topic: Proteomics Chair: Zhiping Liu

16:20-16:45 A two-step framework for inferring direct protein-protein interaction network from AP-MS data Bo Tian, Can Zhao, Feiyang Gu and Zengyou He Dalian University of Technology PaperID: 14
16:45-17:10 A novel hybrid method for protein structure prediction problem on 3D lattice Yuzhen Guo, Fengying Tao and Yong Wang Nanjing University of Aeronautics and Astronatics PaperID: 44
17:10-17:35 An innovative strategy to design scoring functions for near native proteinprotein interface recognition Yongxiao Yang and Xinqi Gong Renmin University of China PaperID: 46

17:35-18:00 *Multiple dimensional feature space for protein interface residue prediction* Gong Xinqi and Cao Tingyi

Renmin University of China PaperID: 47

18:00-20:00 Banquet

Poster Session (Academic Conference Room at 2nd floor 二楼学术报告厅) Edge-network based early-warning signals of Influenza Infection Xiangtian Yu, Tao Zeng and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 5 Analyzing personalized early-warning signals of Type 1 diabetes on microbiome by landscape of dynamical network biomarkers Lu Wang and Jiarui Wu Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 6 Characteristics and evolution of CRISPR-based system in Cyanobacteria Genomes Yan Cai, Jingsong Zhang and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 8 Constrained target controllability-based method for drug target identification in *biological network* Guo Weifeng, Zhang Shaowu, Wei Zegang, Zeng Tao, Wu Fangxiang, Liu Fei, Zhang Jingsong and Chen Luonan Northwestern Polytechnical University PaperID: 9 Reconstructing networks of individual samples based on UMVUE and ensemble learning Yiwei Zhou and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 12 Extended partial correlation for measuring direct associations in networks Juan Zhao and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 13 The key regulators in the browning of white adipocyte Lei Zhang and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 18 Detecting early-warning signals of hepatocellular carcinoma by dynamical network biomarkers Lina Lu, Luonan Chen and Tao Zeng Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 22 Detecting dynamical network biomarker in acute metabolic disease Zhonglin Jiang, Xiaoping Liu and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 32

Diabetes recovery: increased cholesterol metabolism in Roux limb makes glucose rebalance in rat model following Roux-en-Y gastric bypass Meiyi Li, Zhiyuan Liu, Huarong Zhou and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 33 *Recovering spatial-temporal networks based on grouping samples* Ziming Wang, Tao Zeng and Luonan Chen ShanghaiTech University PaperID: 35 Comparative Network Stratification: identifying functional network biomarkers Zhang Chuanchao Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 51 Consistently dysfunctional gene-pairs reveal subtype-specific signatures of non-small cell *lung cancer* Qiangian Shi and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 52

August 21 (Sunday) Technical sessions

06:30-08:30 Beach football match dedicated to 10th anniversary of ISB (Beach before 1st floor of the hotel 宾馆一楼外沙滩)

08:50-10:30 Highlight Session A4 (Academic Conference Room at 2nd floor 二 楼学术报告厅)

Topic: Bioinformatics Chair: Xing-Ming Zhao

08:50-09:15 Applications of Integrative OMICs Approaches to Gene Regulation Studies Jing Qin, Bin Yan, Yaohua Hu, Panwen Wang and Junwen Wang The Chinese University of Hong Kong PaperID: 54
09:15-09:40 Rapid evolutionary turnover underlies conserved lncRNA-genome interactions Qiangfeng Cliff Zhang and Howard Chang Tsinghua University

PaperID: 56

09:40-10:05 *Comparison and Interpretation of Taxonomical Structure of Bacterial Communities in Two Types of Lakes on Yun-Gui plateau of China*

Maozhen Han, Yanhai Gong, Chunyu Zhou, Junqian Zhang, Zhi Wang and Kang Ning

Huazhong University of Science and Technology PaperID: 59

08:50-10:30 Session B4 (Haina-Yunsong Conference Room at 1st floor 一楼海 纳云松会议厅)

Topic: Biomarker Chair: Shihua Zhang

08:50-09:15 DNB Landscape: Local DNB and Its Application on Intestinal Luminal Dynamics Associated with Aging in Mice

Fang Duan, Shinji Fukuda, Luonan Chen and Kazuyuki Aihara University of Tokyo, Japan PaperID: 26

09:15-09:40 An Efficient Differentially Expressed Genes Selection Method for Nervous Diseases

Xueting Huo, Han Zhang and Xin Su Nankai University PaperID: 49

09:40-10:05 Detecting critical state before phase transition of complex biological systems by hidden Markov model

Pei Chen, Rui Liu and Luonan Chen

South China University of Technology PaperID: 53 10:05-10:30 Network-regularized Sparse Logistic Regression Models for Clinical Risk Prediction and Biomarker Discovery Wenwen Min, Juan Liu and Shihua Zhang Wuhan University PaperID: 57

10:30-10:50 Coffee break

10:50-12:30 Session A5 (Academic Conference Room at 2nd floor 二楼学术报告厅)

Topic: Bioinformatics Chair: Minghua Deng

10:50-11:15 A 2-Approximation Scheme for Sorting Signed Permutations by Reversals, *Transpositions, Transreversals, and BlockInterchanges*

Fanchang Hao, Melvin Zhang and Hon Wai Leong Shandong University of Political Science and Law PaperID: 16

11:15-11:40 *Improvement of Phylogenetic Method to Analyze Compositional Heterogeneity*

Fei Guo Tianjin University PaperID: 20

11:40-12:05 *Estimating Phred scores of Illumina DNA base calls using logistic regression models*

Bo Wang, Sheng Zhang, Lin Wan and Lei Li

Academy of Mathematics and Systems Science, Chinese Academy of Sciences PaperID: 30

12:05-12:30 *Multi-level structure for constructing avian influenza virus system based on granular computing*

Yang Li, Qi-Hao Liang, Meng-Meng Sun, Xu-Qing Tang and Ping Zhu Jiangnan University PaperID: 31

10:50-12:30 Session B5 (Haina-Yunsong Conference Room at 1st floor 一楼海 纳云松会议厅)

Topic: Systems Biology Chair: Kang Ning

10:50-11:15 Modeling and Analysis of the Delta-Notch Dependent Boundary Formation in the Drosophila Large Intestine

Fei Liu, Deshun Sun and Hiroshi Matsuno Harbin Institute of Technology

11:15-11:40 Gliomas cell fate decisions mediated by Dll1-Jag1-Fringe in Notch1 signal pathway Xiaofei Shi and Ruiqi Wang

Shanghai University PaperID: 29

11:40-12:05 DMAK: A Curated Pan-Cancer DNA Methylation Annotation

Knowledgebase

Binhua Tang Hohai University PaperID: 39

12:05-12:30 *MeRIPWalker: Identification of RNA methylation driven genes with restart random walk*

Songyao Zhang, Shaowu Zhang, Lian Liu, Jia Meng and Yuifei Huang Northwestern Polytechnical University PaperID: 50

12:30-13:30 Lunch

14:00-15:40 Plenary Session P2 (Academic Conference Room at 2nd floor 二 楼学术报告厅)

Chair: Xianghong Jasmine Zhou

14:00-14:50 The genomics study of urinary system tumor: from reading to writing Zhiming Cai

Shenzhen Second People's Hospital (The First Affiliated Hospital of Shenzhen University), China

14:50-15:40 Expansion of biological pathways by integrative Genomics Jun Liu Harvard University, USA

15:40-16:20 Coffee break & Poster Session

16:20-18:00 Session A6 (Academic Conference Room at 2nd floor 二楼学术报告厅)

Topic: Bioinformatics

Chair: Lei Li

16:20-16:45 Efficiently mining compact k-mers by inverted projection

Jingsong Zhang, Bingqiang Liu, Tao Zeng, Xiangtian Yu, Weifeng Guo and Luonan Chen

Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences PaperID: 7

16:45-17:10 Learning a structural and functional representation for gene expressions: To robustly dissect complex cancer phenotypes Yanbo Wang, Quan Liu, Shan Huang and Bo Yuan Shanghai Jiao Tong University PaperID: 24
17:10-17:35 Recovering hidden diagonal structures via non-negative matrix factorization with multiple constraints Xi Yang, Guoqiang Han, Hongmin Cai and Yan Song South China University of Technology PaperID: 41
17:35-18:00 Deep network architecture for protein-protein interaction interface residue pair prediction Zhenni Zhao and Xinqi Gong Renmin University of China PaperID: 42

16:20-18:00 Session B6 (Haina-Yunsong Conference Room at 1st floor 一楼海 纳云松会议厅)

Topic: Systems Biology Chair: Jingzhi Lei

16:20-16:45 Detecting early-warning signals for influenza A pandemic based on protein dynamical network biomarkers Jie Gao, Kang Wang and Tao Ding Jiangnan University PaperID: 10 **16:45-17:10** Escape rate in biochemical reactions and stochastic integrals Jinhua Wang, Quan Liu, Yanbo Wang and Bo Yuan Shanghai Jiao Tong University PaperID: 21 17:10-17:35 Coordination of Plant Primary Metabolism Studied with a Constraint-based Metabolic Model of C3 Mesophyll Cell Zhuo Wang, Linying Lu, Lin Liu and Jian Li Shanghai Jiao Tong University PaperID: 23 17:35-18:00 Optimal control of tumor treatment with oncolytic virus and MEK inhibitor Yongmei Su University of Science and Technology Beijing PaperID: 34

18:30-20:00 Dinner

August 22 (Monday) Social Program

08:00-13:00 Half-day excursion in Weihai. Departure at Hotel Lobby

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

Book of Abstracts

Plenary Sessions

Structure-Function Mapping of 3D Human Genome

Xianghong Jasmine Zhou University of California Los Angeles, USA

Three-dimensional (3D) genome structures vary from cell to cell even in an isogenic sample. Unlike protein structures, genome structures are highly plastic, posing a significant challenge for structure-function mapping. Recently we developed an approach to comprehensively identify 3D chromatin clusters that each occurs frequently across a population of genome structures, either deconvoluted from ensemble-averaged Hi-C data or from a collection of single-cell Hi-C data. Applying our method to a population of genome structures (at the macrodomain resolution) of lymphoblastoid cells, we identify an atlas of stable inter-chromosomal chromatin clusters. A large number of these clusters are enriched in binding of specific regulatory factors and are therefore defined as 'Regulatory Communities.' We reveal two major factors, centromere clustering and transcription factor binding, which significantly stabilize such communities. Finally, we show that the regulatory communities differ substantially from cell to cell, indicating that expression variability could be impacted by genome structures.

Transcript-Based Differential Expression Analysis for Population RNA-Seq Data

Tao Jiang University of California - Riverside, USA Tsinghua University, China

Differential transcript expression (DTE) analysis on population data without predefined conditions is critical to many biological or clinical studies. For example, it can be used to discover biomarkers to classify cancer samples into previously unknown subtypes such that better diagnosis and therapy methods can be developed for the subtypes. Although several DTE tools have been published, these tools either assume binary conditions in the input population or require the number of conditions to be given. In this work, we propose a novel DTE analysis algorithm, called SDEAP, that estimates the number of conditions directly from the input samples using Dirichlet mixture models and discovers alternative splicing events using a new graph modular decomposition algorithm. By taking advantage of the above technical improvement, SDEAP was able to outperform the other DTE analysis methods in our extensive experiments on simulated data and real data with qPCR validation. The prediction of SDEAP also allowed us to classify the samples of cancer subtypes and cell-cycle phases more accurately.

This is joint work with Ei-Wen Yang at UCR/UCLA.

The genomics study of urinary system tumor: from reading to writing

Zhiming Cai Shenzhen Second People's Hospital (The First Affiliated Hospital of Shenzhen University), China

From the discovery of cancer gene to the "read" (reading and elucidating) of human genome, the human beings went through a long journey. Now the United States launched projects of "writing gene" (editing gene) in organisms. During the past 27 years, our team experienced the whole process of the reading of urinary system tumor gene, and also reached to the international level in terms of gene writing.

In the aspects of gene reading: we successively carried out the single gene, multiple gene and genomics studies in children Wilms tumor, renal, bladder and urinary system tumor by using the PCR and next-generation sequencing technology; We found eight chromatin remodeling bladder cancer gene, and this was commented as the "chromatin remodeling - tumor suppressor theory" for cancer propagation; Besides, we elucidated the mechanisms of renal cancer development and transformation; In addition, we discovered the highest mutation gene (Driver gene - TERT) in bladder cancer; then we developed the new discovered genes as diagnostic markers to develop new technology in early cancer screening using urine.

In the aspects of gene writing: we constructed the artificial gene circuit and screened the mutant peptide in renal and bladder cancer with synthetic biology technology, and also remodeled the engineered immune cells and applied the NY-ESO1, PDL-1 and Car-T in immune cells in accurate treatment; In addition, we established the world's largest bladder cancer multiple genomics database of yellow race, and also undertook the task of international research alliance of bladder cancer and kidney cancer genome project on behalf of China; Furthermore, the project "the synthetic biology device intervention in bladder cancer research" won the national 973 plan project and the leader Cai was awarded as "chief scientist". The research results of Cai were published in Nature, Nature Genetics, Cell and other well-known publications. Besides, Cai received high praise from 《Urinary system tumor genomics research》 and was awarded the first prize for Guangdong province natural sciences in 2015.

Expansion of biological pathways by integrative Genomics

Jun Liu Harvard University, USA

The number of publicly available gene expression datasets has been growing dramatically. Various methods had been proposed to predict gene co-expression by integrating the publicly available datasets. These methods assume that the genes in the query gene set are homogeneously correlated and consider no gene-specific correlation tendencies, no background intra-experimental correlations, and no quality variations of different experiments. We propose a two-step algorithm called CLIC (CLustering by Inferred Co-expression) based on a coherent Bayesian model to overcome these limitations. CLIC first employs a Bayesian partition model with feature selection to partition the gene set into disjoint co-expression modules (CEMs), simultaneously assigning posterior probability of selection to each dataset. In the second step, CLIC expands each CEM by scanning the whole reference genome for candidate genes that were not in the input gene

set but co-expressed with the genes in this CEM. CLIC is capable of integrating over thousands of gene expression datasets to achieve much higher coexpression prediction accuracy compared to traditional co-expression methods. Application of CLIC to ~1000 annotated human pathways and ~6000 poorly characterized human genes reveals new components of some well-studied pathways and provides strong functional predictions for some poorly characterized genes. We validated the predicted association between protein C7orf55 and ATP synthase assembly using CRISPR knock-out assays.

Based on the joint work with Yang Li and the Vamsi Mootha lab.

Parallel Sessions

Paper ID: 1

Molecular dynamics simulation reveals how phosphorylation of tyrosine 26 of phosphoglycerate mutase 1 upregulates glycolysis and promotes tumour growth

Yan Wang and Guanyu Wang

South University of Science and Technology of China, China

Phosphoglycerate mutase 1 (PGAM1) is a crucial enzyme in glycolysis. Its 26th residue tyrosine(Y26) is frequently found phosphorylated in many cancer cells, indicating its power in promoting tumour growth. Besides, the phosphorylation of Y26 of PGAM1 was implied to affect the binding of the 2,3-bisphosphoglycerate (2,3-BPG) molecule. Till now, structure-function relationships of Y26 phosphorylation PGAM1 enzyme have not been studied at the atomic resolution. To this end, the exact effect of the phosphorylation of Y26 of PGAM1 was investigated by employing the molecular dynamic (MD) simulation combined with binding free energy calculations. After Y26 phosphorylation, additional salt bridge interactions were found of this residue. Such interactions stabilized the loop region this residue located. In such stable state, another residue of this loop region, Ser22 played the role in stabilization the 2,3-BGP molecule. The knowledge of the phosphorylation of Y26 of PGAM1 may deepen the understanding of Y26 phosphorylation related tumour growth.

Paper ID: 2 **Identifying network-based biomarkers of complex diseases from high-throughput data** *Zhi-Ping Liu Shandong University, China*

In this talk, we will review the main available computational methods of identifying biomarkers of complex diseases from high-throughput data. The emerging omics techniques provide powerful alternatives to measure thousands of molecules in cells in parallel manners. The generated genomic, transcriptomic, proteomic, metabolomic and phenomic data provide comprehensive molecular and cellular information for detecting critical signals served as biomarkers by classifying disease phenotypic states. Networks are often employed to organize these profiles in the identification of biomarkers to deal with complex diseases in diagnosis, prognosis and therapy as well as mechanism deciphering from systematic perspectives. Here, we summarize some representative network-based bioinformatics methods in order to highlight the importance of computational strategies in biomarker discovery.

Paper ID: 3

An Improved Feature-Based Approach to Predict Effective Drug Combinations *Qian Xu, Yi Xiong, Hao Dai, Kotni Meena, Qin Xu, Hongyu Ou and Dongqing Wei* Combinatorial therapy is a promising strategy for combating complex disorders, such as cancers and AIDS. In spite of the increasing number of drug combinations in use, many of them were found in the clinic by experience rather than designation. In this study, a new computational method is proposed to predict effective drug combinations by integrating biological, chemical and pharmacological information. We collected a set of 352 pairs of effective drug combinations from Drug Combination Database. Then, a 732 dimensional feature vector involving biological, chemical and pharmaceutical information was constructed for each drug combination to describe its properties. To avoid over-fitting, Minimum Redundancy Maximum Relevance was performed to extract useful ones by removing redundant subsets. After feature selection, three different machine learning algorithms were applied to build the drug combination prediction models. Our results demonstrated that the model based on stochastic gradient boosting yield out the best performance. Furthermore, feature patterns of therapy were found to have powerful ability to discriminate effective drug combinations from non-effective ones. By comparing and combining various features, it is shown that the enriched features occurred frequently in golden positive samples not only help predict novel drug combinations, but also offer a unique perspective on the implicit mechanism of combinational drugs.

Paper ID: 4

The network biomarker for complex diseases

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

Based on the accumulation of biomedical big data, the precision medicine for complex diseases is in quick development. As a core component to bridge the foundational exploration to clinical application, the biomarkers and their identification urgently require advanced hypothesis, theory and approach to simultaneously promote the pathogen discrimination and biological interpretation of molecule markers. Dissimilar to conventional "node" biomarkers, i.e. using individual and separate genes as molecule markers for the diagnosis or prognosis of complex diseases; we have proposed new "edge" biomarkers, i.e. adopting multiple molecules and their (expression) associations to be the network biomarkers of complex diseases.

Network biomarkers, on one hand, introduces the constraints of molecule association to improve the biological interpretation; on the other hand, also designs new quantitative measurements of networks to enhance the numerical discrimination. Based on the experiments on a large number of disease datasets, network biomarkers are actually capable to extract novel pathogen differential information from omics data, i.e. discover the differential correlations among " darkened " non-differentially expressed molecules; furthermore, the quantification of such differential correlation on "small-sample-large-dimension" high-throughput data can promote the accuracy of disease prediction. Particularly, the network biomarkers can describe new individual differences, so that, it can help to investigate the disease subtypes in the viewpoint of heterogeneity and supply new insights on the personalized disease prevention, prediction and treatment.

Edge-network based early-warning signals of Influenza Infection

Xiangtian Yu, Tao Zeng and Luonan Chen

Key Laboratory of Systems Biology, Innovation Center for Cell Signaling Network, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China

To study the early diagnosis of infectious diseases, the analysis of host-based gene expression shows a great potential, especially for influenza prediction. We have developed a novel edge-based network, i.e. edge-network, to detect early signals of diseases (i.e. influenza) by identifying the edge-biomarkers quantified with warning score from dynamical network biomarkers. Specifically, we show that the stochastic network of a biological system can be described by the integration of its node-network and its edge-network in an accurate manner, where the novel edge-network is a representation of second-order statistics gene expression profiles and the traditional node-network is a representation of the first-order statistics of the noisy data. The edge-network analysis has been carried on the gene expressions of human volunteers within live influenza experiment on either influenza A/Brisbane/59/2007 (H1N1) or A/Wisconsin/67/2005 (H3N2)), which have the samples of peripheral blood transcriptome at every 8 hours for 7 days. The edge-biomarkers (with 265 genes for H1N1 and also 265 genes for H3N2) have been identified in the edge-networks corresponding to symptomatic adults, which can be used to predict the subsequent outcomes of influenza infection. Especially, these edge-biomarkers completely cover all 22 influenza genes we found previously. Actually, we correctly predict the final infection outcome of each individual just with the expression data before his/her clinic symptom by these edge-biomarkers, in which the prediction accuracy as AUC achieves 100% for H3N2 and 82.22% for H1N1 respectively. Furthermore, we demonstrate the superiority of our method by cross-marker test between H3N2 and H1N1. The prediction accuracy still achieves 100% for H3N2 with H1N1-specific edge-biomarkers, and 81.48% for H1N1 with H3N2-specific edge-biomarkers. Thus, the edge-network analysis not only opens a new way to understand the cross-species pathogenesis at a network level, but also provides a powerful tool to make the early prediction of complex diseases.

Paper ID: 6

Analyzing personalized early-warning signals of Type 1 diabetes on microbiome by landscape of dynamical network biomarkers

Lu Wang and Jiarui Wu

Shanghai Institutes for Biological Sciences & University of Science and Technology of China, China

Motivation: Type 1 diabetes (T1D) is a disorder disease resulting from an immune-mediated dysfunction of pancreatic beta cells in genetically predisposed individuals. The intestinal microbiome represents a symbiotic ecological community that influences human development and

function, and has great impacts upon the development of T1D. We hypothesized that dynamic changes in microbial community indicate the progression of T1D, and there is a set of signatures of composition perturbation which can be used to predict T1D. Methods: As known to us, species interactions can be inferred from a single system monitored over several time points. Therefore, we inferred a series of co-occurrence networks based on personalized temporal intestinal microbiome data. And a new landscape model of dynamical network biomarker (DNB) is developed and applied to detect the critical phase transitions of T1D individuals, where a DNB represents a subnetwork of the whole microbial community network Results: Based on the landscape model of DNB, we computed a warning index which measures the amount of microbe variation, indicating an imminent deterioration before the critical transition occurs. Through our method, we detected a few DNB microbes, and found that the patterns of microbe like B. fragilis could directly impact the intestinal mucosa, cause the inflammation, and thereby increase the risk of type 1 diabetes. These alterations in the microbiome preceding T1D onset may provide novel biomarkers and therapeutic targets for prevention or treatment of T1D.

Paper ID: 7

Efficiently mining compact k-mers by inverted projection

Jingsong Zhang, Bingqiang Liu, Tao Zeng, Xiangtian Yu, Weifeng Guo and Luonan Chen Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China School of Mathematics, Shandong University, China School of Automation, Northwestern Polytechnical University, China

K-mer counting in biological sequences is a typical sequence analysis problem with broad applications. Existing k-mer counting methods count the full set of k-mers satisfying a specified minimum support threshold in a sequence database. However, such conventional counting poses a great challenge at spawning numerous redundant or meaningless k-mers, and even is inherently costly in runtime when the support threshold is low. Inspired by the application of the upper-closure property of sequential patterns, we hold an assumption that k-mers would have a compact form like closed k-mers, and propose CloKmer, an efficient algorithm to find the complete set of "closed k-mers", which have significant implications in both topology and biology. Particularly, CloKmer utilizes an effective data structure called inverted index to project the original biological sequences. This algorithm combines a pattern-growth scheme, three search space pruning techniques, and an equivalence-class-based upper-closure checking fashion to perform the closed k-mer mining. We experimentally evaluate various aspects of CloKmer by using both DNA and protein sequence datasets. Our results demonstrate that CloKmer significantly outperforms the previous methods in terms of effectiveness and efficiency, but with similar requirement on memory usage. Furthermore, our analyses also show that the closed k-mers are superior to the general k-mers in terms of both compactness and biological significance of patterns.

Characteristics and evolution of CRISPR-based system in Cyanobacteria Genomes Yan Cai, Jingsong Zhang and Luonan Chen Institute of Biochemistry and Cell Biology, China

CRISPR-based system widely exists in bacteria and archaea genomes, and works as a defensive system to protect prokaryotic cells from foreign DNA attack. Cyanobacteria is one of the earliest lives on earth, and is a phylum of bacteria that obtain their energy through photosynthesis. Now, more and more Cyanobacteria genomes have been sequenced and published, and CRISPR structures also are found in some of these genomes. According to the conserved sequence structure of the CRISPR-based system, blasting the whole genomic sequences, the location and the enviroment of CRISPR in multiple genomes can be obtained. From the relevant information as CRISPR number, location, structure motif arrangement in each Cyanobacteria species, the characteristics of CRISPR-based system in Cyanobacteria genomes can be extracted, and the evolutionary relationship between CRISPR structure and different Cyanobacteria species will also reveal more interesting hints on the transfer of CRISPR and evolution story of cyanobacteria.

Paper ID: 9

Constrained target controllability-based method for drug target identification in biological network

Guo Weifeng, Zhang Shaowu, Wei Zegang, Zeng Tao, Wu Fangxiang, Liu Fei, Zhang Jingsong and Chen Luonan

Northwestern Polytechnical University, China Shanghai Institutes for Biological Science, Chinese Academy Science, China University of Saskatchewan, Saskatoon

Analyzing complex networks from the viewpoint of controllability will lead us to a deeper understanding about complex systems. Furthermore, it is of theoretical interest and practical significance to study how to control the system by applying some perturbations to driver nodes. However, most of existing studies often focus on how to obtain the state transitions of the system by selecting driver nodes in the whole network. In practice, to control a complex network, one may know not only the set of nodes which need to be controlled (i.e. target nodes) but also the set of nodes to which only control signals can be applied (i.e. constrained control nodes). It is more practical and necessary to control target nodes by applying control signals to constrained control nodes, prompting us to study the constrained target controllability of complex networks: the ability to drive any state of target nodes to their desirable state by applying control signals to the driver nodes from the set of constrained control nodes with the minimum cost. To efficiently investigate CTC of complex networks, here we present a novel graphic-theoretic algorithm, called CTCA, which integrates the target controllable subsystems, to appropriate the minimum number of driver nodes. Applications to real bimolecular networks identified efficiently drug targets for their phenotype transitions and validated our method.

Detecting early-warning signals for influenza A pandemic based on protein dynamical network biomarkers

Jie Gao, Kang Wang and Tao Ding Jiangnan University, China

There is critical phenomenon for the outbreak of influenza A in which the influenza A comes from a relatively stable state to the state of outbreak. In this paper, different states of influenza A in the method of dynamical network biomarkers (DNB) are studied. Through establishing DNB of influenza A virus protein and using the nature of DNB, early warning signals for influenza A outbreak are detected and a composite index is ultimately obtained. The composite index varies along with the state of pandemic influenza virus from a relatively steady state to critical state before outbreak and then to the outbreak state. When the composite index stays at low value (<1)steady state, it means influenza A is normally in the relatively steady stage. When the composite index of a certain year increases by more than 0.8 relative to the previous year, and it is less than 1, and it increases sharply and reaches a peak and larger than 1 in next year, it means the year is normally in the critical state before outbreak and the next year is normally in the outbreak state. Therefore, we can predict and identify whether a certain year is in the critical state before influenza A outbreak or outbreak state by observing the variation of index value. This indicates the composite index can provide a reliable and significant warning information to detect the stage of influenza A, which will be significantly meaningful for the warning and prevention of influenza A pandemic.

Paper ID: 11

Mathematical model reveals competitive roles of autophagy in inhibiting progression and promoting growth in pancreatic cancers

Jinzhi Lei

Tsinghua University, China

Inhibiting autophagy recently has been considered as an effective cancer therapy. In pancreatic cancer, autophagy inhibition attenuates the late-grade tumor growth, however, is also found to accelerate progression in some cases. Experimental data suggest effect of autophagy may depend on p53 status. Here we present a mathematical model consisting of four-stage pancreatic cancer progression and their interplay with autophagy, p53, and DNA damage. The model is able to reconcile several existing seemingly contradictory experimental observations, making specific predictions on different role of p53 in each stage during cancer progression. Through modeling, p53 status at PanIN-3 is found to be a critical factor in modulating autophagy effect in progression, and more desirable function of autophagy may be achieved by regulating DNA damage, suggesting a new treatment strategy by combining NAC and autophagy therapy. Overall, the competition between the two opposite functions of autophagy mediated by p53 determines the cancer dynamics and growth.

Reconstructing networks of individual samples based on UMVUE and ensemble learning *Yiwei Zhou and Luonan Chen*

Shanghai Institutes for Biological Sciences Chinese Academy of Sciences, ShanghaiTech University, China

In general, reconstructing a network or quantifying an edge requires multiple samples, e.g., at least three samples. However, multiple biological or medicine samples are not available for many cases. In this work, we show that for a number of samples following the same margin distribution, we can reconstruct individual network of each sample, i.e., infer the network by only one sample. Specifically, assuming a joint bivariate distribution on samples, the edge strength or the correlation between two variables can be estimated by one sample. In particular, we established a statistic to estimate the correlation or edge strength between two variables in the network of each sample. It can be proved that our statistic is an UMVUE (uniformly minimum-variance unbiased estimator) to the correlation. In some cases, the accuracy of this statistic will be up to 100%. By applying ensemble learning, we can further improve the accuracy the estimation on the reconstructed network. Both the simulated data and real biological data validate the effectiveness and efficiency of our method on individual sample networks.

Paper ID: 13

Extended partial correlation for measuring direct associations in networks Juan Zhao and Luonan Chen

shanghai institutes for biological sciences, CAS, China

Measuring direct dependencies between variables is an important task in data analysis, in particular for reconstructing various types of networks and causal relations in science and engineering. One of the most widely used criteria is partial correlation, but it may fail to detect the associations when strong associations occur in variables. In this work, we propose a new measure, "Extended partial correlation" based on information theory, which not only can overcome the problem of partial correlation but also retains the quantification properties of partial correlation. Specifically, we first defined extended partial correlation to measure direct dependencies between variables and then derived its relations partial correlation and conditional mutual information. Finally, we used a number of simulated data as benchmark examples to numerically demonstrate the extended partial correlation features and further real gene expression data from Escherichia coli and yeast to reconstruct gene regulatory networks.

Paper ID: 14

A two-step framework for inferring direct protein-protein interaction network from AP-MS data

Bo Tian, Can Zhao, Feiyang Gu and Zengyou He School of Software, Dalian University of Technology, Dalian, China Affinity purification-mass spectrometry (AP-MS) has been widely used for generating bait-prey data sets so as to identify underlying protein-protein interactions and protein complexes. However, the AP-MS data sets in terms of bait-prey pairs are highly noisy, where candidate pairs contain many false positives. Recently, numerous computational methods have been developed to identify genuine interactions from AP-MS data sets. However, most of these methods aim at removing false positives that contain contaminants, ignoring the distinction between direct interactions and indirect interactions. In this paper, we present an initialization-and-refinement framework for inferring direct PPI networks from AP-MS data, in which an initial network is first generated with existing scoring methods and then a refined network is constructed by the application of indirect association removal methods. Experimental results on several real AP-MS data sets show that our method is capable of identifying more direct interactions than traditional scoring methods.

Paper ID: 15

Accelerated Parallel Algorithm for Gene Network Reverse Engineering Jing He, Zhou Zhou, Michael Reed and Andrea Califano Columbia University, USA

The Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe) represents one of the most effective tools to reconstruct gene regulatory networks from large-scale molecular profile datasets. However, previous implementations suffer from the requirement for intensive computing resources, limitations of sample sizes, or inefficiency of execution. One way to address this challenge is to solve the problem of massively repeated mathematical computation, which is the essential task, using a GPU computing framework. However, the re-design of a serial algorithm into a parallelized one requires fine optimization efforts based on a deep understanding of both hardware structure and algorithm details. Herein, we present an accelerated parallel implementation of ARACNe based on GPU (GPU-ARACNe). By taking advantage of multi-level parallelism and the Compute Unified Device Architecture (CUDA) parallel kernel-call library, GPU-ARACNe successfully parallelizes a serial algorithm and simplifies the user experience from multi-step operations to one step. Using public datasets on comparable hardware configurations, we showed that GPU-ARACNe is faster than previous implementations. We showed that the networks inferred from GPU-ARACNe are valid. Two versions of GPU-ARACNe implementations are provided. GPU-ARACNe-V1 is readily runnable on any GPU with computing capability 3.5 or higher. GPU-ARACNe-V2 runs on any GPU of computing capability less than 3.5, such as Amazon Web Service GPU instances. Given that the previous versions of ARACNe are extremely demanding, either in computational time or investment in hardware, GPU-ARACNe is extremely valuable for researchers lacking computational resources to build complex regulatory networks from hundreds of gene expression profiles. In addition, our optimization of adaptive partitioning for Mutual Information estimation using GPU provides lessons that are applicable to other domains. The source code and a small dataset are publicly available in Bitbucket: https://moonhj00@bitbucket.org/moonhj00/gpu-aracne.git.

A 2-Approximation Scheme for Sorting Signed Permutations by Reversals, Transpositions, Transreversals, and BlockInterchanges

Fanchang Hao, Melvin Zhang and Hon Wai Leong Shandong University of Political Science and Law, China National University of Singapore, Singapore

We consider the problem of sorting signed permutations by reversals, transpositions, transreversals, and block-interchanges. In 2003, He and Chen [12] gave a 2-approximation algorithm for sorting signed permutations by reversals and block-interchanges. In 2005, Hartman and Sharan [11] gave a 1.5 approximation algorithm for sorting signed permutations by transreversals and transpositions. We give a 2-approximation scheme for sorting signed permutations by reversals, transpositions, transreversals, and block-interchanges. We call it the GSB (Genome Sorting by Bridges) scheme. Our result extends the work of He and Chen [12] by allowing transreversals (generalized reversals) and that of Hartman and Sharan [11] by allowing block-interchanges (generalized transpositions). We prove this result by introducing three bridge structures in the breakpoint graph, namely, the L-bridge, T-bridge, and X-bridge and show that they model "good" reversals, transpositions / transreversals, and block-interchanges, respectively. We show that we can always find at least one of these three bridges in any breakpoint graph, thus giving an upper bound on the number of operations needed. We prove a lower bound on the distance and use it to show that GSB has a 2-approximation ratio. An O(n3) algorithm called GSB-I that is based on the GSB approximation scheme presented in this paper has since recently been published by Yu, Hao and Leong in [17]

Paper ID: 18 **The key regulators in the browning of white adipocyte** *Lei Zhang and Luonan Chen SIBS, China*

Adipose tissue in mammals, can be divided into two types according to the different functions and colors, i.e., the white adipose tissue (WAT) and the brown adipose tissue (BAT). Their distribution in body is different, and they are also different in cellular morphology and molecular markers. Conventional wisdom holds that the BAT is abundant in newborns' body. However, as living organisms grow older, the BAT in their body would disappear. But in recent years, people found by the position emission tomography (PET) that the BAT also exists in adults' body and has the physiological function of thermogenesis. Obesity is a kind of chronic non-communicable disease which is caused by excessive accumulation of energy in WAT. Obesity can influence the level of insulin, blood sugar and blood fat. And these problems would trigger the type 2 diabetes, hypertension, non-alcoholic fatty liver disease, atherosclerosis and many other metabolic diseases. From the energy balancing perspective, to increase the heat production and energy expenditure through activating the BAT is a good way to lose weight. However, the relations between the BAT and WAT remain largely unknown. In this work, we try to find the mechanism of the browning of WAT by comparing the gene expression profiles of two kinds of adipose tissues. We found 166 up-regulated genes and 28 down-regulated genes in brown adipose tissue. By constructing the network of those genes, we identified the key regulatory factors of the lipid metabolism, and also

analyzed the functional roles, which are strongly related to the metabolic disease, such as obesity and diabetes.

Paper ID: 19

Drug target prediction by multi-view low rank embedding

Menglan Cai and Limin Li Xi'an Jiaotong University, China

Drug repositioning has been a key problem in drug development ,and heterogeneous data sources are used to predict drug-target interactions by different approaches. However, most of studies focus on single representations of drugs or proteins. It has been shown that multi-view representations of drugs and proteins can strengthen the prediction ability. For example, a drug can be represented by its chemical structure, or by its chemical response in different cells. A protein can be represented by its sequence, or by its gene expression values in different cells. The docking of drugs and proteins based on their structure can be considered as one view (structural view), and the chemical performance of them based on gene expression and drug response can be considered as another view (chemical view). In this work, we first propose a single-view approach of SLRE based on low rank embedding for an arbitrary view, and then extend it to a multi-view approach of MLRE which could integrate both views. Our experiments show that our methods perform significantly better than baselines methods including single-view methods and multi-view methods.

Paper ID: 20 **Improvement of Phylogenetic Method to Analyze Compositional Heterogeneity** *Fei Guo Tianjin University, China*

Phylogenetic analysis is a key way to understand current research in the biological processes and detect theory in evolution of natural selection. We extract the biological data via modeling features, and then compare these characteristics to study the biological evolution between species. The evolutionary relationship between species is generally reflected in the form of phylogenetic trees. Many methods for constructing the phylogenetic tree, are based on the optimization criteria. Here, we use maximum likelihood and bayesian inference method to establish phylogenetic trees; while a certain extent, multi-chain Markov chain Monte Carlo sampling can be used to select optimal phylogenetic tree, resolving local optimum problem. The correlation model of phylogenetic analysis assumes that phylogenetic tree are built on homogeneous data, however there exists a large deviation in the presence of heterogeneous data. We use conscious detection to solve compositional heterogeneity. Our method is evaluated on two sets of experimental data, a group of bacterial 16S ribosomal RNA gene data, and a group of genetic data with five homologous species. Our method can obtain accurate phylogenetic tree on the homologous data, and also detect the compositional heterogeneity of experimental data. We provide an efficient method to enhance the accuracy of generated phylogenetic tree.

Escape rate in biochemical reactions and stochastic integrals

Jinhua Wang, Quan Liu, Yanbo Wang and Bo Yuan Shanghai Jiao Tong University, China

In this article, we consider an essential problem in an evolutionary biological system with its dynamic behavior described by a stochastic differential equation - escape from a metastable state. We study the limiting case where the mass approaches zero and derive the classical escape rate formula for a diffusion process with a state-dependent friction coefficient using different stochastic integrals. For the first time, to our knowledge, we show that the anti-Ito integral is the naturally one leading to the classical result.

Paper ID: 22

Detecting early-warning signals of hepatocellular carcinoma by dynamical network biomarkers

Lina Lu, Luonan Chen and Tao Zeng Shanghai Institutes for Biological Sciences, China

Hepatocellular carcinoma (HCC) is a complex disease, with multi-step process stages from preneoplastic lesions-including cirrhosis, low grade dysplasia (LGD) and high grade dysplasia (HGD) to HCC including early HCC and advanced HCC. The molecular mechanisms associated with pathogenesis of HCC are still unclear and there is no effective pro-carcinogenic biomarker to predict the pre-disease state of HCC.

In this work, we identified the critical transition of hepatocarcinogenesis and its early-warning signals based on the gene expression profiles (GSE6764) during the stepwise carcinogenic process by dynamical network biomarker (DNB) as a new leading biomolecular networks used for early diagnosis of the disease. We found that some upcoming sudden network changes appear around the LGD period during the HCC progression, after which a state transition or drastic deterioration occurs. The identified dynamical network biomarkers can be used as the early-warning signals of HCC. Furthermore, we demonstrated that the leading biomolecular networks of DNBs are strongly related to the initiation of HCC, and can provide several biological insights into the molecular mechanism of pathogenesis of HCC. The functional analysis e.g., GO and KEGG, on DNBs and other experimental data from literatures can well validate our computational results.

Paper ID: 23

Coordination of Plant Primary Metabolism Studied with a Constraint-based Metabolic Model of C3 Mesophyll Cell

Zhuo Wang, Linying Lu, Lin Liu and Jian Li Shanghai Jiao Tong University, China Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, China Most of the current efforts in plant metabolic engineering focused on either individual enzymes or a few enzymes in a particular pathway without considering the potential interactions between metabolic pathways. However, the complex interaction between different pathways in plant metabolic network requires consideration of the complex interdependence between pathways to maximize the chance of success of engineering efforts. Constraint-based models, which do not require detailed kinetic parameters for the involved reactions, provide an opportunity to study the potential interaction between pathways in plant metabolic network. In this study, we developed a constraint-based model of C3 plant primary metabolism (C3PMM) and used it to study the coordination between metabolic pathways. We first demonstrated that there is substantial coordination of mesophyll primary metabolism at transcriptome and metabolism levels under changed CO2 concentrations. Secondly, we demonstrate that maximizing CO2 uptake is a plausible objective function for metabolism of a typical C3 mesophyll cell. Finally, C3PMM predicted a decrease in nitrate assimilation flux and increase in ammonium assimilation flux when CO2 concentration increases, suggesting a potential mechanistic linkage between differences in the response of ecosystems differing in nitrogen source to elevated CO2. In conclusion, this study demonstrated the potential application of using constraint-based modeling in guiding plant metabolic engineering and studying response of plant metabolism under climate change.

Paper ID: 24

Learning a structural and functional representation for gene expressions: To robustly dissect complex cancer phenotypes

Yanbo Wang, Quan Liu, Shan Huang and Bo Yuan Shanghai Jiao Tong University, China Department of Neurosurgery Shanghai 5th People's Hospital and Shanghai Huashan Hospital Fudan University School of Medicine, China

Cancer is a heterogeneous disease, thus one of the central problems is how to dissect the resulting complex phenotypes in terms of their biological building blocks. Computationally, this is to represent and interpret high dimensional observations through a structural and conceptual abstraction into the most influential determinants underlying the problem. The working hypothesis of this report is to consider gene interaction to be largely responsible for the manifestation of complex cancer phenotypes, thus where the representation is to be conceptualized. Here we report a representation learning strategy combined with regularizations, in which gene expressions are described in terms of a regularized product of meta-genes and their expression levels. The meta-genes are constrained by gene interactions thus representing their original topological contexts. The expression levels are supervised by their conditional dependencies among the observations thus providing a cluster-specific constraint. We obtain both of these structural constraints using a node-based graphical model. Our representation allows the selection of more influential modules, thus implicating their possible roles in neoplastic transformations. The modules discovered are either shared or specify for different types or stages of human cancers, all of which are consistent with literature and biology. Finally, we validate our representation strategy by its robust recognitions of various cancer phenotypes comparing with various classical methods.

Network-based method to understanding the therapeutic effect of ShexiangBaoxin Pill on cardiovascular diseases

Jing Zhao Logistical Engineering University, China

ShexiangBaoxin Pill (SBP), a Traditional Chinese Medicine formula to treat cardiovascular disease (CVD), is universally used in clinical practice in China. However, its action mechanism is not well interpreted yet, due to its complex components and targets. In this paper, approaches of network pharmacology are utilized to further reveal the action mechanism that SBP exerts on CVD from perspective of protein-protein interactions and pathways. Meanwhile, the fact that SBP's plasma absorbed compounds play main therapeutic role in targeting CVD, with respect to its full compound set, is confirmed. Finally, gene expression data from microarray experiments is performed to verify our computational results. This study may facilitate our understanding of anti-CVD effect of SBP from perspective of network pharmacology.

Paper ID: 26

DNB Landscape: Local DNB and Its Application on Intestinal Luminal Dynamics Associated with Aging in Mice

Fang Duan, Shinji Fukuda, Luonan Chen and Kazuyuki Aihara The University of Tokyo, Japan Keio University, Japan Chinese Academy of Sciences, China

The dynamical network biomarkers (DNBs) method can detect early-warning signals of genetic networks and serve as predisease biomarkers. The DNBs are based on bifurcation theory, which highlights the alterations of a network attractor to others. In this work, we present a DNB landscape method under the framework of DNB to investigate the intestinal luminal dynamics associated with aging in mice. In the traditional DNB, the composite index (CI) of a network is obtained by calculating the normalized indices of a cluster so called the leading network. The leading network with strong fluctuations and strong correlations holds a high DNB index score, which indicates that the network is closer to a tipping point between attractors. In the proposed method, the local CI of each node is calculated, instead of the CI of the overall leading network. Each local CI is calculated based on a reference network. The standard deviations of the node, correlations between the node and the first order neighbors, and the correlations between the node and nodes outside the first order neighbors are three parts of the local CI. Therefore, the DNB landscape method can map the landscape of DNB across the entire network. Meanwhile, the proposed method could avoid an open question of how to choose the leading network. We applied the DNB landscape method to intestinal microbiome and metabolome data of specific pathogen-free (SPF) C57BL/6 mice. The common reference networks were estimated by significance tests of all time points of the dataset. The local DNB index indicated a critical point at

60 weeks, which is related to the aging process of SPF mice. The selected nodes, so called the leading nodes, with high local CI were mainly chemical components of amino acids in plasma. The results were compatible with other studies of aging and demonstrated that the proposed method could detect the critical change of biochemical networks of SPF mice.

Paper ID: 27

Modeling and Analysis of the Delta-Notch Dependent Boundary Formation in the Drosophila Large Intestine

Fei Liu, Deshun Sun and Hiroshi Matsuno Harbin Institute of Technology, China Yamaguchi University, Japan

The boundary formation in the Drosophila large intestine is widely studied as an important biological problem. It has been shown that the Delta-Notch signaling pathway plays one of important roles in the formation of boundary cells. In this paper, we propose a mathematical model for the Delta-Notch dependent boundary formation in the Drosophila large intestine with the purpose of better interpreting the related experimental findings of this biological phenomenon. To achieve this, we not only perform stability anlysis on the model from a theoretical point of view, but also numerically analyze the influences of perturbation of some key parameters on the boundary formation, and the sensitivity of the expression rate of Delta and Notch and obtain the value intervals for resulting in different patterns in both normal and mutant conditions. With all these work, we can assure that our model can better interpret the biological phenomena related to the boundary formation in the Drosophila large intestine.

Paper ID: 28

Dynamics evolutions of HIV infection on treatment with combinatorial drugs *Lu Xiaojing, Liu Yanwei and Wang Ruiqi Department of Mathematics, Shanghai University, Shanghai, China*

At present, the highly active antiretroviral treatment is recognized as the most effective therapy to treat and control HIV/AIDS. Furthermore, it is still very concerned that why combination drugs is better than the single drugs. In recent decades, there are already some researches about the multiple combinatorial perturbation theory. However, these observations mainly rely heavily on the dynamics of transient characteristics. Due to the sensitivity of the transient dynamics to the initial state, above methods have large amount of calculation. In this paper, we propose a method only rely on the steady state, i.e. the combinatorial perturbation research based on the theory of the branch to depict why drug combinations is better than single drug treatment. In this work, we first proposed a basic mathematical model of HIV infection without drug resistence to verify that the monotherapy with reverse transcriptase inhibitors could present rapid suppression of HIV replication in the short-term. Furthermore, we modified the basic model by taking into the immune response mechanism and drug resistence. Using a bifurcation-based method of combinatorial perturbation we analyze the synergistic regulation action of combination drugs on treatment of

HIV infection. As a result, our analysis provided some theoretical supports of combination drugs therapy for AIDS from the point of view of dynamics theory based on bifurcation.

Paper ID: 29

Gliomas cell fate decisions mediated by Dll1-Jag1-Fringe in Notch1 signal pathway

Shi Xiaofei and Wang Ruiqi Shanghai University, China

The Notch family of proteins plays a vital role in determining cell fates, such as proliferation, differentiation, and apoptosis. It has been proved that Notch1 and its ligands, Dll1 and Jag1, are overexpressed in many glioma cell lines and primary human gliomas. In this article, we devise a specific theoretical framework that incorporates Dll1, Jag1 and fringe in Notch1 signaling circuit to explore the functional role of these proteins in gliomas cells in the tumorigenesis and progression of human gliomas, accordingly to provide some proof of theory for treatment of human gliomas. We mainly want to know the role of fringe in the whole process, futher to predict the method for the treatment of gliomas. We find that the change of notch1, its ligands and fringe will change the steady state in the system.

Paper ID: 30

Estimating Phred scores of Illumina DNA base calls using logistic regression models *Bo Wang, Sheng Zhang, Lin Wan and Lei Li*

Academy of Mathematics and Systems Science, CAS, China

Phred quality scores, defined as the error probabilities of each base, are very helpful in downstream DNA analysis such as SNP detection and DNA assembly. Thus a valid model to define Phred scores is indispensable for any base-calling software that produces sequences from raw signals. Recently, we developed a software 3Dec for Illumina sequencing platforms (https://github.com/flishwnag/3dec), which reduces base-calling errors by 44-69% at a given mapping rate compared to the existing ones. However, the model to predict quality scores was not fully investigated yet. In this study, we estimated the quality scores for 3Dec based on logistic regression models. We selected 76 features as the predictive variables in the initial model. They include features derived from the signals after correction for color-, cyclic- and spatial-crosstalk, the current positions, the two most likely nucleotide bases of current positions and the called bases of neighbor positions. The error labels of the called bases were obtained by mapping the sequences to the reference genome. Next the initial model was reduced by three methods: backward elimination with AIC, backward elimination with BIC, and L1-regularized learning method. We used two measures to assess the methods: 1. the AUC of the ROC curve, which represents the discrimination power of a method between high- and low- quality bases; 2. the consistency between empirical quality scores with the predicted ones.We trained the models using features and error labels from ~300,000 bases in the first tile of the BlindCall Hiseq2000 dataset (http://www.cbcb.umd.edu/~hcorrada/secgen/), and tested the models on sequences from the third tile. While the three methods achieved similarly high AUC scores (0.9141, 0.9161, and 0.9173,

respectively), L1-regularized logistic model calibrates the consistency between empirical and expected scores better than the other two methods. And its computation is much faster. Thus itwould be applied in a later version of 3Dec to improve the prediction of quality scores. The above modelling not only defines valid Phred scores, but also provides insights to the error mechanism of the sequencing technology. We found several interesting patterns. First, a nucleotide 'T' after a 'G' was more likely to be miscalled. Second, nucleotides were more likely to be miscalled as the previous bases if the previous ones were not 'G'. This suggested that the phasing effect of bases after 'G' was somewhat different from those after other nucleotide types.

Paper ID: 31

Multi-level structure for constructing avian influenza virus system based on granular computing

Yang Li, Qi-Hao Liang, Meng-Meng Sun, Xu-Qing Tang and Ping Zhu Jiangnan University, China

Exploring the genetic structure of influenza viruses attracts the attention in the field of molecular ecology and medical genetics, whose epidemics causes morbidity and mortality worldwide annually. The rapid variation in RNA strand and changes of protein structure make the subtyping identification low accuracy and difficult to develop effective drugs and vaccine. This paper constructs the hierarchical structure of avian influenza virus system considering both HA and NA segments. An optimization model was established based on the fuzzy hierarchical evaluation index to determine the rational granularity of the virus system for exploring the intrinsic relationship among the subtypes. Thus, an algorithm was presented to extract the optimal structure and the multi-level structure could be constructed by repeating applying the proposed algorithm. Furthermore, to reduce the systematic and computational complexity, the granular signatures of virus system are identified based on the coarse-grained idea and then its performance will be evaluated through a designed classifier. The results showed that the proposed method could identify the effective virus signatures and improve subtype identification. Furthermore, the obtained virus signatures could approximate and reflect the whole system. Once a new molecular virus is detected, the prevention and treatment measures can follow what were applied in the viral signature.

Paper ID: 32

Detecting dynamical network biomarker in acute metabolic disease Zhonglin Jiang, Xiaoping Liu and Luonan Chen Shanghai Institutes for Biological Sciences Chinese Academy of Sciences, China

Metabolic diseases, such as non-alcoholic fatty liver disease(NAFLD) and obesity, have become the most common and emerging form of chronic human disease. However, the regulatory processes of these diseases are still not clear due to complex interplay of many internal and external factors. With the rapid advances of highthroughput technologies, such as microarray, next-generation sequencing, genomics and proteomics, big data generated from the technologies provide unprecedented opportunities to study the molecular mechanisms of the metabolic diseases from a system-wide perspective. In this work, we combined dynamical network biomarker (DNB) and biological network to detect the early signal of these complex diseases and analyze the potential disease associated pathway based on a time series RNA-Seq data from the model mouse with NAFLD.

Paper ID: 33

Diabetes recovery: increased cholesterol metabolism in Roux limb makes glucose rebalance in rat model following Roux-en-Y gastric bypass

Meiyi Li, Zhiyuan Liu, Huarong Zhou and Luonan Chen SIBS, China

Recently it has been found that Roux-en-Y gastric bypass surgery can help unexpectedly to improve obesity and hyperglycemia. However, the underlying mechanisms or roles have been little known. Here, we performed RYGB procedures in Goto-Kakizaki (GKS) rats with identical pancreatic duodenal limb lengths. After surgery 12 weeks, blood glucose levels of the treated GK rats are similar to the Wistar rats' and better than the sham-operated/ pair-fed GK (GK-PF-Sham) rats'. Interestingly, comparing with the whole genomic expression profiles of these three different groups of rats, we found that although the treated GKS rats recovered from diabetes they were showed unusually different with the normal Wistar rats. To uncover the roles of RYGB in improvements of hyperglycemia, we grouped and characterized different change patterns according to the directions of gene expression change to their normal levels in Wistar rats. Then, it was unexpectedly found that in the new-changed pattern many genes involved in CHOL synthesis play a main role in rebalance of glucose.

Paper ID: 34 **Optimal control of tumor treatment with oncolytic virus and MEK inhibitor** *Yongmei Su Su University of Science and Technology Beijing, China*

This paper gives a new mathematical model of tumor therapy with oncolytic virus and MEK inhibitor. Stable analysis was given. Because inhibitors (MEK) can not only lead to more oncolytic virus into cancer cells, but also can limit the replication of the virus, in order to provide the best dosage of MEK inhibitors and balance the positive and negative effect of the inhibitors, we put forward an optimal control problem of the inhibitor. The optimal strategies are given by theory and simulation.

Paper ID: 35 **Recovering spatial-temporal networks based on grouping samples** Ziming Wang, Tao Zeng and Luonan Chen ShanghaiTech University, China Key Laboratory of Systems Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China

Estimating the time-dependent or stage-wise networks is of great importance to our understanding of biological processes. Rapid advance of modern technologies made it possible to measure multiple omics data (e.g. proteome) in time series. These data can be used to construct the time-dependent or group-specific networks on different individuals. However, due to the strong noise of high-throughput data and the small size of samples, the direct estimation of such networks may lead to suspicious or even incorrect results. In this work, we extended the previous model of inferring network structures to a sophisticated form, in which we can consider the similarity within spatial or temporal sample-groups, i.e., called group-based time-varying Lasso (gtLasso). We showed that gtLasso can guarantee high recovery of each time-dependent network corresponding to a particular group at a specific time period even with limited time points. Provided that there are sufficient number of samples in each group, we can infer the accurate network structure. We also applied gtLasso to integrative analysis of the proteome and metabolism data on individuals with RYGB operation. The preliminary results demonstrate the effectiveness and efficiency of our method.

Paper ID: 36

BDB: biopanning data bank

Bifang He, Guoshi Chai, Yaocong Duan and Jian Huang University of Electronic Science and Technology of China, China

The BDB database (http://immunet.cn/bdb) is an update of the MimoDB database, which was previously described in the 2012 Nucleic Acids Research Database issue. The rebranded name BDB is short for Biopanning Data Bank, which aims to be a portal for biopanning results of the combinatorial peptide library. Last updated in July 2015, BDB contains 2904 sets of biopanning data collected from 1322 peer-reviewed papers. It contains 25 786 peptide sequences, 1704 targets, 492 known templates, 447 peptide libraries and 310 crystal structures of target-template or target-peptide complexes. All data stored in BDB were revisited, and information on peptide affinity, measurement method and procedures was added for 2298 peptides from 411 sets of biopanning data from 246 published papers. In addition, a more professional and user-friendly web interface was implemented, a more detailed help system was designed, and a new on-the-fly data visualization tool and a series of tools for data analysis were integrated. With these new data and tools made available, we expect that the BDB database would become a major resource for scholars using phage display, with improved utility for biopanning and related scientific communities.

Paper ID: 37 **Data science – part of integrated agriculture solution** *Le Lv and Jingdong Liu Monsanto company, China* Monsanto is committed to bringing a broad range of solutions to help nourish our growing world. Through programs and partnerships, we collaborate with farmers, researchers, nonprofit organizations, universities and others to help tackle some of the world's biggest challenges. This talk contains general information about how we levrage data science to provide integrated solutions for farmers, as well as using systems biology gene network approaches to accelerate agriculture product development.

Paper ID: 38

Network clustering analysis using Mixture Exponential-family Random Graph Models and its application in Genetic Interaction

Yishu Wang, Huaying Fang, Dejie Yang, Minghua Deng and Hongyu Zhao School of Mathematical Sciences, Qingdao University, China School of Mathematical Sciences & Center for Quantitative Biology, Peking University, China School of Information, Central University of Finance and Economics, China LMAM, School of Mathematical Sciences, Peking University, China Department of Biostatistics, Yale School of Public Health, USA

Motivation: Epistatic miniarrary prole (EMAP) studies have enabled the mapping of large-scale genetic interaction networks and generated large amounts of data in model organisms. It provide an incredible set of molecular tools and advanced technologies that should be efficiently understanding the relationship between the genotypes and phenotypes of individuals. However, the network information gained from EMAP cannot be fully exploited using the traditional statistical network models. Because the genetic network are always heterogeneous, for example, the network structure features for one subset of nodes are different from those of the left nodes. Exponential-family random graph models (ERGMs) are a family of statistical models, which provides a principled and flexible way to describe the structural features (e.g. the density, centrality and assortativity) of an observed network. However, the single ERGM is not enough to capture this heterogeneity of networks. In this paper, we consider a mixture EGRM (mixEGRM) networks, which model a network consisted of several communities, where the subnetwork of each community is describe by a single EGRM. Results: EM algorithm is a classical method to solve the mixture problem, hower, it will be very slow when the data size is huge in the numerous application. We adopt an efficient novel online graph clustering algorithm to classify the class of the graph nodes and estimate the ERGM parameters for the mixERGM. In simulation studies, the mixERGM outperforms the role analysis for the network cluster in which the mixture of exponential-family random graph model is developed for many ego-network according to their roles. One genetic interaction network of yeast and Two real social networks show the wide potential application of the mixERGM.

Paper ID: 39 DMAK: A Curated Pan-Cancer DNA Methylation Annotation Knowledgebase Binhua Tang

Hohai University, China Shanghai Jiao Tong University, China

Pan-cancer analysis can identify cell- and tissue-specific genomic loci and regions with underlying biological functions. Here we present an online curated DNA Methylation Annotation Knowledgebase, DMAK, which includes the pan-cancer analysis results for differentially-methylated loci and regions by the Reduced Representation Bisulfite Sequencing profiling technology. DMAK contains three modules of curated information and analysis results on 688,445 CpG sites across 19 cancer and embryonic stem cell lines from ENCODE, and further analysis of survival associations with clinical sources retrieved from TCGA. The knowledgebase covers all identified differentially-methylated CpG sites and regions of interest, further annotated genomic information, together with tumor suppressor genes information and calculated methylation level. DMAK provides meaningful clues for deriving functional association network and related clinical association results based on protein-coding genes, including tumor suppressor genes, identified from differentially methylated regions of interest. Thus DMAK constitutes a comprehensive reference source for the current epigenetic research and clinical study.

Paper ID: 40

Meta-analysis for Feature Selection Based on Gene Co-expression Network *Xue Jiang, Han Zhang and Xiongwen Quan*

Nankai University, China

Predicting disease gene by analyzing gene expression data has become a popular theme, thus efficient feature selection and dimensionality reduction have become fundamental tools for processing high-dimensional gene expression data. Most existing feature selection methods deal with gene expression data directly, resulting in high false positive rates in feature subsets. To our knowledge, no or little effort has been devoted to methods that consider the interactions between genes and the dynamic processes of gene co-expression in co-expression network for feature selection. In this study, an efficient meta-analysis method for feature selection based on gene co-expression network(MFSN) was presented. In MFSN, we defined the concepts of differential co-expression of genes in co-expression network, then calculated the gene differential co-expression value according to the topological structure changes between different phase-specific networks, finally conducted meta-analysis of gene differential co-expression based on the rank product method. Experimental results demonstrated the effectiveness of MFSN and its superior performance over other popular feature selection methods through a set of real-world gene expression data sets.

Paper ID: 41

Recovering hidden diagonal structures via non-negative matrix factorization with multiple constraints

Xi Yang, Guoqiang Han, Hongmin Cai and Yan Song School of Computer Science & Engineering, South China University of Technology, China Revealing data with intrinsically diagonal block structures is particularly useful for analyzing groups of highly correlated variables. Earlier research based on non-negative matrix factorization (NMF) has been shown to be effective at representing such data by decomposing the observed data into two factors, where one factor is considered to be the feature and the other the expansion coefficients from a linear algebra perspective. If the data are sampled from multiple independent subspaces, the coefficient factor would possess a diagonal structure under an ideal matrix decomposition. However, the standard NMF method and its variants have not been reported to exploit this type of data via direct estimation. To address this issue, a non-negative matrix factorization with multiple constraints model is proposed in this paper. The constraints include an 11 -norm on the feature matrix and a (TV,11)-norm on its loading matrix. The proposed model is shown to be capable of efficiently recovering diagonal block structures hidden in observed samples. An efficient numerical algorithm using the alternating direction method of multipliers (ADMM) model is proposed for optimizing the new model. The proposed method is shown to perform robustly and effectively for both simulated data and real biological data. The superior performance of the proposed method compared with two benchmark models demonstrates its applicability in recovering diagonal block structures from observed samples.

Paper ID: 42

Deep network architecture for protein-protein interaction interface residue pair prediction

Zhenni Zhao and Xinqi Gong

Institute of Mathematical sciences, Renmin University of China, China

Motivation: Proteins usually fulfill their biological functions through interacting with other proteins. Although some methods have been developed to predict the binding sites of a monomer protein, which is not enough for prediction of the interaction of two monomer proteins. The correct predictions of across interface residue pairs from two monomer proteins is still an open question and have great significance for practical life science experiment applications. We hope to build an interface residue pairs prediction method suitable for those applications.

Results: Here, we developed a novel deep network architecture called multi-layered Long-Short Term Memory (LSTM) network approach for protein interface residue pairs prediction. First, we characterized the surface residues by nine descriptors and employed these features to discriminate between interface residue pair and non-interface residue pair. Second, we used twice thresholds to select residue pairs which are more likely to be interface residue pairs. What's more, this step increases the proportion of interface residue pairs and reduces the influence of imbalanced data. Third, we built deep neural network architectures based on Long-Short Term Memory algorithm to organize and refine the prediction of interface residue pairs by employing features which are easily obtainable. We trained the deep networks on dimers at unbound state in the international protein-protein interaction prediction contest benchmark version 3.0. The updated data set in the versions 4.0 and 5.0 were used as validation set and test set respectively. For our best model, the accuracy rate is over 62% when we chose top 2‰ pairs of every dimer in test set as interface residue pairs, which will be very helpful for understanding of protein-protein interaction mechanisms and for guidance of biological experiments.

Paper ID: 43 **Modeling of the dynamics of hematopoiesis system in cyclical thrombocytopenia** *Jinzhi Lei and Changjing Zhuge Tsinghua University, China Beijing University of Technology, China*

Cyclical thrombocytopenia (CT) is a rare hematological dissease characterized by periodic oscillations only in the platelets while stem cells, neutrophils and erythocytes remains sustained levels. To understand the mechanisms of this phenomenon, we established mathematical models to describe the dynamics by delay differential equations. Through the models, we simulates the platelet oscillations and sustained states of other blood cells and find out the key component that causes the oscillation in platelets. By numerical and theoretical analysis, we also study the effects of various physiological processes (such as the production of blood cells, the proliferation rates, death rates of cells, etc) on the dynamics of the hematopoiesis system.

Paper ID: 44

A novel hybrid method for protein structure prediction problem on 3D lattice Yuzhen Guo, Fengying Tao and Yong Wang Nanjing University of Aeronautics and Astronatics, China Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China

Predicting protein structure from its amino acid sequence is a prominent problem in computational biology. In this paper, we established an optimization model on the basis of hydrophobic-hydrophilic model on 3D square lattice to translate the biological problem into mathematical problem. Particle Swarm Optimization was extended to solve discrete model. In order to avoid premature problem, modified Tabu Search Strategy was proposed. In addition, based on the characteristic of native protein structure, Pull Strategy was designed to accelerate the convergence of algorithm. The novel hybrid method (TPPSO2) combining Particle Swarm Optimization and optimal s-trategies can predict protein structure more accurately and efficiently on 3D square lattice by comparing with other algorithms.

Paper ID: 45

Bayesian network model for identification of pathways by integrating protein interaction with genetic interaction data

Changhe Fu, Guangxu Jin, Su Deng and Zu-Guo Yu Shenyang Normal University, China Center of Systems Biology and Bioinformatics, Wake Forest School of Medicine, Winston-Salem, NC, USA School of Mathematics and Computational Science, Xiangtan University, Xiangtan, China Many methods have applied to infer biological pathways using molecular interaction data at proteomic and genetic levels. Both of the interaction data provide signaling and functional insights into a molecular biosystem and are helpful for reconstruction of pathway structures complementarily. Here, we apply Bayesian network to integrate protein these two types of interaction data to discover details of pathway structure. We modeled the pathway network as Bayesian network and utilize this model to infer pathways for an endoplasmic reticulum gene set with genetic interactions quantitatively measured by the phenotypes of the unfolded protein response of endoplasmic reticulum. The protein interaction data was derived from the BioGRID database. By comparing N-Glycan biosynthesis pathway identified from our approach with that from APN (activity pathway networks), we found that our approach can overcome the weakness of the APN, that is, certain edges are inexplicable. Moreover, according to underlying protein interaction network, we defined a simple scoring function that only adopts gene pairwise score to avoid the balance difficulty in the APN. Using the effective stochastic simulation algorithm, the performance of our proposed method is significantly high.

Paper ID: 46

An innovative strategy to design scoring functions for near native protein-protein interface recognition

Yongxiao Yang and Xinqi Gong

Institute for Mathematical Sciences, Renmin University of China, China

Motivation: Theoretical prediction of the structure of protein-protein complexes remains an outstanding question. One of the difficulties is that how to recognize the near native protein-protein interfaces from a large number of docking decoys efficiently. A variety of scoring functions to overcome this difficulty have been designed according to the interface properties such as shape complementarity, electrostatics and desolvation. Some of these scoring functions also used the predicted interface information, the discriminative ability strongly depended on the reliability of predicted interface information. Here, we propose an innovative strategy for designing new scoring functions which use the probabilities of all the surface residue pairs (composed of two residues from different partners) as interface ones, the strategy doesn't request the high accuracy of predicted interface as a near native one are powerful for near native interface recognition. The strategy used to construct new scoring functions has a certain significance to the future research on protein-protein interactions

Paper ID: 47 **Multiple dimensional feature space for protein interface residue prediction** *Gong Xinqi and Cao Tingyi Renmin University of China, China*

Protein-protein interacts through specic interface residues to execute biological functions. Correctly understanding the mechanisms of interface recognition and predicting the interface residues are important for many aspects of life science studies. Here, we report a novel architecture to study and predict protein interface residues. In our method, a multiple dimensional feature space was built on some meaningful features. Then we grided the space and put all the surface residues into the grids according to their feature values. Interestingly, interface residues were found to prefer some grids clustered together. We obtained excellent prediction result on a public and veried data benchmark. Our approach not only opens up a new thought direction for interface residue prediction, but also will help to understand protein-protein interactions more deeply.

Paper ID: 48

Computational probing protein-protein interactions targeting small molecules

Yongcui Wang, Shilong Chen, Naiyang Deng and Yong Wang Northwest Institute of Plateau Biology, CAS, China College of Science, China Agricultural University, China National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, CAS, China

Motivation: With the booming of interactome studies, a lot of interactions can be measured in a high throughput way and large scale datasets are available. It is becoming apparent that many different types of interactions can be potential drug targets. Compared with inhibition of a single protein, inhibition of protein-protein interaction (PPI) is promising to improve the specificity with fewer adverse side-effects. Also it greatly broadens the drug target search space, which makes the drug target discovery difficult. Computational methods are highly desired to efficiently provide candidates for further experiments and hold the promise to greatly accelerate the discovery of novel drug targets. Results: Here, we propose a machine learning method to predict PPI targets in a genomic-wide scale. Specifically, we develop a computational method, named as PrePPItar, to Predict PPIs as drug targets by uncovering the potential associations between drugs and PPIs. First, we survey the databases and manually construct a gold-standard positive dataset for drug and PPI interactions. This effort leads to a dataset with 227 associations among 63 PPIs and 113 FDA-approved drugs and allows us to build models to learn the association rules from the data. Second, we characterize drugs by profiling in chemical structure, drug ATC-code annotation, and side-effect space and represent PPI similarity by a symmetrical S-kernel based on protein amino acid sequence. Then the drugs and PPIs are correlated by Kronecker product kernel. Finally, a support vector machine (SVM), is trained to predict novel associations between drugs and PPIs. We validate our PrePPItar method on the well-established gold-standard dataset by cross-validation. We find that all chemical structure, drug ATC-code, and side-effect information are predictive for PPI target. Moreover, we can increase the PPI target prediction coverage by integrating multiple data sources. Follow-up database search and pathway analysis indicate that our new predictions are worthy of future experimental validation. Conclusion: In conclusion, PrePPItar can serve as a useful tool for PPI target discovery and provides a general heterogeneous data integrative framework.

Paper ID: 49 **An Efficient Differentially Expressed Genes Selection Method for Nervous Diseases** *Xueting Huo, Han Zhang and Xin Su Nankai University, China*

High-throughput data mining has become a hot direction of big data mining. Existing biology-related feature ranking methods mainly focus on statistical and annotation information. In this experiment, a gene selection technique was proposed to select differential expression genes from microarray data accurately for the prediction of nervous diseases. In this regard, a hybrid scoring criterion was designed as a linear combination of two parts of scores determined in the q-value based multi-target regression and spectral clustering based on a modified genes co-expression network. In the first part, we considered about both global margin information and locality manifold information at the same time, and in the second part we adopted spectral clustering algorithms based on graph theory. to search for similar genes in genes co-expression network. Empirical experiments demonstrate the effectiveness of this algorithm.

Paper ID: 50

MeRIPWalker: Identification of RNA methylation driven genes with restart random walk

Songyao Zhang, Shaowu Zhang, Lian Liu, Jia Meng and Yuifei Huang Northwestern Polytechnical University, China Department of Biological Sciences, HRINU, SUERI, Xi'an Jiaotong-Liverpool University, USA Department of Electrical and Computer Engineering, University of Texas at San, San Antonio

As the most prevalent mammalian mRNA epigenetic modifications, N6-methyladenosine (m6A) is identified to influence gene expression and link to some significant physiological processes, i.e., obesity, synaptic signalling, cancer, plant development, sperm development and so on. Nevertheless, what roles m6A plays and what factors m6A regulates in these physiological processes are still mystery. Therefore, to reveal the function of RNA methylation, we focus on RNA methylation driven genes which are genes with RNA hyper or hypomethylation under certain stimulation, i.e., disease or certain gene knockout, compared to normal RNA methylation state. Considering biological replicates of MeRIP-seq data and functional interconnection among genes in certain physiological process, we propose a novel algorithm, MeRIPWalker, combining PPI network, to build consensus RNA methylation driven gene network from m6A modification data under different cell status. MeRIPWalker detected genes with RNA hyper or hypomethylation in each biological replicates set and identified RNA methylation driven genes by the joint frequency of genes with differential RNA methylation and their candidate genes. The candidate genes were prioritized by the Random Walk with Restart algorithm in a protein-protein interaction network and filtered by their topological and biological significance. We applied MeRIPWalker to build 4 networks respectively from FTO, METTL3, METTL14 and WTAP gene knock out data which are transcriptome wide m6A methylation data sequenced by MeRIP-seq. Consistent across the 4 networks, the top enriched GO biological process terms have smaller p-values than those of exomePeak algorithm, indicating that the RNA methylation driven genes identified by MeRIPWalker have greater biological significance. Another comparative experiment with

VarWalker algorithm shows MeRIPWalker is faster and more robust. Functional enrichment analysis of the 4 networks reveals several significant biological progress, cellular pathways and diseases, especially cancer, related to m6A, which is consistent with previous studies and the enriched genes are deduced to be significant factors regulated by m6A.

Paper ID: 51 **Comparative Network Stratification: identifying functional network biomarkers** *Zhang Chuanchao*

Chinese Academy of Sciences, China

A major challenge of bioinformatics in the era of precision medicine is to identify the molecular biomarkers for complex diseases. It is a general expectation that these biomarkers or signatures have not only strong discrimination ability, but also readable interpretations in a biological sense. Generally, the conventional expression-based or network-based methods mainly capture differential genes or differential networks as biomarkers, however, such biomarkers only focus on phenotypic discrimination and usually have less biological or functional interpretation. Meanwhile, the conventional function-based methods could consider the biomarkers corresponding to certain biological functions or pathways, but ignore the differential information of genes, i.e., disregard the active degree of particular genes involved in particular functions, thereby resulting in less discriminative ability on phenotypes. Hence, it is strongly demanded to develop elaborate computational methods to directly identify functional network biomarkers with both discriminative power on disease states and readable interpretation on biological functions. In this paper, we present a new computational framework based on an integer programming model, named as Comparative Network Stratification (CNS), to extract functional or interpretable network biomarkers, which are of strongly discriminative power on disease states and also readable interpretation on biological functions. In addition, CNS can not only recognize the pathogen biological functions disregarded by traditional Expression-based /Network-based methods, but also uncover the active network-structuress underlying such dysregulated functions underestimated by traditional Function-based methods. To validate the effectiveness, we have compared CNS with five state-of-the-art methods, i.e. GSVA, Pathifier, stSVM, frSVM and AEP on four datasets of different complex diseases. The results show that CNS can enhance the discriminative power of network biomarkers, and further provide biologically interpretable information or disease pathogenic mechanism of these biomarkers. A case study on T1D demonstrates that CNS can identify many dysfunctional genes and networks previously disregarded by conventional approaches. Therefore, CNS is actually a powerful bioinformatics tool, which can identify functional or interpretable network biomarkers with both discriminative power on disease states and readable interpretation on biological functions.

Paper ID: 52 Consistently dysfunctional gene-pairs reveal subtype-specific signatures of non-small cell lung cancer Qianqian Shi and Luonan Chen

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

A complex disease is generally caused by genetic alterations or disorder of biological regulation. Systematic identification of dysfunctional gene-pairs can shed light on the mechanisms underlying complex diseases, and provide crucial information to develop efficient biomarkers or therapies. In this paper, we proposed a novel approach to detect consistently dysfunctional gene-pairs for non-small cell lung cancer, from normal to disease status in heterogeneous disease datasets. In particular, we determined reliable disordered regulations between normal and disease states by identifying the consistent differential topological network structures. To validate the effectiveness of our method, two subtypes of non-small cell lung cancer data were used to identify the consistently dysfunctional gene-pairs from normal to cancer onset, and the results well agree with the known information as well as the experiments, thereby implying predictive power of the method. The comparison with other approaches also indicated the superiority of our method, and actually those identified gene-pairs can be used as new edgetic targets for network drug design. In contrast to molecular biomarkers, we also show that the consistently dysfunctional gene-pairs can be applied for predicting patient survial for both subtypes of lung cancers.

Paper ID: 53

Detecting critical state before phase transition of complex biological systems by hidden Markov model

Pei Chen, Rui Liu and Luonan Chen South China University of Technology, China Shanghai Institutes for Biological Sciences, China

For most biological processes, the dynamics of phase transition is generally composed of three stages, i.e., before-transition state, pre-transition state, and after-transition state. Unlike traditional detection of an after-transition state in which there are clear phenomena, it is usually a challenge to identify the pre-transition state during the progression of a biological system just before the occurrence of a critical transition, not only because of the high complexity of the biological system, but there may be few clues appearing until the catastrophic phase transition really takes place. In this work, by exploiting the differential network features in dynamics between the before- and pre-transition states, we present a hidden-Markov-model (HMM) based computational method to identify the pre-transition state and elucidate the essential mechanisms during the critical transition at the network level. Specifically, by considering the network variation in dynamics and regarding that the pre-disease state is the end or shift-point of a stationary Markov process, an inconsistence score is proposed to measure the probability that a system is in consistency with the normal state. As validation, this approach is applied to detect the upcoming critical transition of complex systems based on both the dataset generated from a simulated network and the rich information provided by high-throughput microarray data. The effectiveness of our method has been demonstrated by the identification of the pre-disease states for two real datasets. Both functional and pathway enrichment analyses validate the computational results.

Applications of Integrative OMICs Approaches to Gene Regulation Studies Jing Qin, Bin Yan, Yaohua Hu, Panwen Wang and Junwen Wang The Chinese University of Hong Kong, China Shenzhen Institutes of Advanced Technology, China Shenzhen University, China Mayo Clinic

The applications of OMICs technologies, such as whole genome/exome sequencing, ChIP sequencing and RNA sequencing, have been widely used in biological and medical research. These OMICs data provide an overview of living cells at DNA, RNA and protein levels. To better understand the complexity of the biological process, particularly the underlying gene regulatory networks, we need powerful bioinformatics tools to integrate these OMICs data. This article reviews the popular OMICs technologies, OMICs data integration strategies, and the bioinformatics tools used for multi-dimensional data integration. We will highlight the advantages of these methods for elucidating molecular basis of biological regulatory mechanisms.

Paper ID: 56

Rapid evolutionary turnover underlies conserved lncRNA–genome interactions Qiangfeng Cliff Zhang and Howard Chang Tsinghua University, China Stanford University, USA

Many long noncoding RNAs (lncRNAs) can regulate chromatin states, but the evolutionary origin and dynamics driving lncRNA-genome interactions are unclear. We adapted an integrative strategy that identifies lncRNA orthologs in different species despite limited sequence similarity, which is applicable to mammalian and insect lncRNAs. Analysis of the roX lncRNAs, which are essential for dosage compensation of the single X chromosome in Drosophila males, revealed 47 new roX orthologs in diverse Drosophilid species across ~ 40 million years of evo- lution. Genetic rescue by roX orthologs and engineered synthetic lncRNAs showed that altering the number of focal, repetitive RNA structures determines roX ortholog function. Genomic occupancy maps of roX RNAs in four species revealed conserved targeting of X chromosome neighborhoods but rapid turnover of individual binding sites. Many new roX-binding sites evolved from DNA encoding a pre-existing RNA splicing signal, effectively linking dosage compensation to transcribed genes. Thus, dynamic change in lncRNAs and their genomic targets underlies conserved and essential lncRNA-genome interactions.

Paper ID: 57

Network-regularized Sparse Logistic Regression Models for Clinical Risk Prediction and Biomarker Discovery

Wenwen Min, Juan Liu and Shihua Zhang

Wuhan University, China Academy of Mathematics and Systems Science, CAS, China

Paper ID: 58

Integrative analysis for identifying joint modular patterns of gene-expression and drug-response data

Jinyu Chen and Shihua Zhang

Academy of Mathematics and Systems Science, CAS, China

Motivation: The underlying relationship between genomic factors and the response of diverse cancer drugs still remains unclear. A number of studies showed that the heterogeneous responses to anticancer treatments of patients were partly associated with their specific changes in gene expression and somatic alterations. The emerging large-scale pharmacogenomic data provide us valuable opportunities to improve existing therapies or to guide early-phase clinical trials of compounds under development. However, how to identify the underlying combinatorial patterns among pharmacogenomics data are still a challenging issue. Results: In this study, we adopted a sparse network-regularized partial least square (SNPLS) method to identify joint modular patterns using large-scale pairwise gene-expression and drug-response data. We incorporated a molecular network to the (sparse) partial least square model to improve the module accuracy via a network-based penalty. We first demonstrated the effectiveness of SNPLS using a set of simulation data and compared it with two typical methods. Further, we applied it to gene expression profiles for 13 321 genes and pharmacological profiles for 98 anticancer drugs across 641 cancer cell lines consisting of diverse types of human cancers. We identified 20 gene-drug co-modules, each of which consists of 30 cell lines, 137 genes and 2 drugs on average. The majority of identified co-modules have significantly functional implications and coordinated gene-drug associations. The modular analysis here provided us new insights into the molecular mechanisms of how drugs act and suggested new drug targets for therapy of certain types of cancers.

Comparison and Interpretation of Taxonomical Structure of Bacterial Communities in Two Types of Lakes on Yun-Gui plateau of China

Maozhen Han, Yanhai Gong, Chunyu Zhou, Junqian Zhang, Zhi Wang and Kang Ning, China Huazhong University of Science and Technology, China Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, China Institute of Hydrobiology, Chinese Academy of Sciences, China Institute of Geodesy and Geophysics, Chinese Academy of Sciences, China

Bacterial communities from freshwater lakes are shaped by various factors such as nutrients, pH value, temperature, etc. Their compositions and relative abundances would undergo changes to adapt the changing environments, and in turn could affect the environments of freshwater lakes. Analyses of the freshwater lake's bacterial communities under different environments would be of pivotal importance to monitor the condition of waterbody. In this study, we have collected freshwater samples from two lakes on Yun-Gui plateau of China, Lake Dianchi and Lake Haixihai, and analyzed the bacterial community structures from these samples based on 16S rRNA sequencing. Results have shown that: Firstly, the bacterial community of these samples have very different taxonomical structures, not only between two lakes but also among the intra-groups for samples collected from Dianchi. Secondly, the differences between samples from two lakes are highly associated with the chemical-geographical properties of the two lakes. Thirdly, for samples of Dianchi and Haixihai, analytical results of physicochemical, taxonomical structure and relative abundance of community revealed that extreme physicochemical factors caused by human activities have strongly affected the bacterial ecosystem in Dianchi. These results have clearly indicated the importance of combining biological profiling and chemical-geographical properties for monitoring Chinese plateau freshwater bacterial ecosystem, which could provide clues for Chinese freshwater ecosystem remediation on plateau.

Paper ID: 60 **Minimize the number of features for a biomedical data mining problem** *Fengfeng Zhou Jilin University, China*

Bio-technologies have been rapidly innovated and improved to produce huge amount of biomedical big data. The Precision Medicine model has challenged the informatics technologies for precise and cost-efficient data mining models of disease diagnosis and prognosis. These technologies may produce millions of measurements for a single participant, and each measurement is regarded as a feature of this sample/participant. Our laboratory focuses on finding the minimal number of features, so that a data mining model may be constructed to accurately detect a phenotype. This is a feature selection problem in the area of data mining. We demonstrated that both multi-step refining and constrained strategies may generate satisfying data mining performances (Ge, et al., 2016; Ma, et al., 2016; Zhou, et al., 2015).

Robustness and controllability of complex networks

Xueming Liu and Linqiang Pan Huazhong University of Science and Technology, China

Increasing evidence shows that real-world systems interact with one another via dependency connectivities. Failing connectivities are the mechanism behind the breakdown of interacting complex systems, e.g., blackouts caused by the interdependence of power grids and communication networks. Previous research analyzing the robustness of interdependent networks has been limited to undirected networks. However, most real-world networks are directed, their in-degrees and out-degrees may be correlated, and they are often coupled to one another as interdependent directed networks. To understand the breakdown and robustness of interdependent directed networks, we develop a theoretical framework based on generating functions and percolation theory. We find that the percolation behavior of interdependent directed scale-free networks with and without degree correlations is so complex that two criteria are needed to quantify and compare their robustness: the percolation threshold and the integrated size of the giant component during an entire attack process. Interestingly, we find that the in-degree and out-degree correlations in each network layer increase the robustness of interdependent degree heterogeneous networks that most real networks are, but decrease the robustness of interdependent networks with homogeneous degree distribution and with strong coupling strengths.

This framework can be used to study the robustness of many real interdependent networks. For example, gene regulatory networks, Protein interaction networks and metabolic networks are interdependent with each other. The framework provides a way to understand how the biological system changes under gene deletions and enables development of drug discovery and effective treatments, which helps to control the biological system.

We apply structural controllability to a human signaling network and detect driver nodes, providing a systematic analysis of the role of different proteins in controlling the human signaling network. We find that the proteins in the upstream of the signaling information flow and the low in-degree proteins play a crucial role in controlling the human signaling network. Interestingly, inputting different control signals on the regulators of the cancer-associated genes could cost less than controlling the cancer-associated genes directly in order to control the whole human signaling network in the sense that less drive nodes are needed. This research provides a fresh perspective for controlling the human cell signaling system.