The 11th International Conference on Computational Systems Biology (ISB 2017)

August 18-21, 2017
Shenzhen, China

ISB 2017 Schedule

August 18 Friday	13:00-18:00	Registration (Hotel Lobby)	
	18:00-20:00	Reception	
	20:00-21:30	Board member meeting of ORSC-CSB (Room MM)	
August 19 Saturday	08:30-08:40	Opening Session (Room FH)	
	08:40-10:10	Plenary Session P1 (Room FH)	
	10:10-10:40	Coffee break	
	10:40-12:10	Plenary Session P2 (Room FH)	
	12:30-13:30	Lunch	
	13:00-14:00	Board member meeting of CSBMB Systems Biology Division (Room MM)	
	14:00-15:40	Highlight Session A1 (Room FH)	Session B1 (Room LT)
		Topic: Bioinformatics	Topic: Network Biology
		Paper IDs: 48, 54, 93, 87	Paper IDs: 3, 11, 25, 68, 82
		Chair: Yu Xue	Chair: Bairong Shen
	15:40-16:20	Coffee break & Poster Session	
	16:20-18:00	Highlight Session A2 (Room FH)	Session B2 (Room LT)
		Topic: Systems Biology	Topic: Bioinformatics
		Paper IDs: 91, 90, 46, 79	Paper IDs: 6, 53, 57, 61, 71
		Chair: Tianshou Zhou	Chair: Jiguang Wang
	18:30-20:00	Banquet	

Room FH: Feng-Huang Conference Room at 2nd floor (二楼凤凰阁)
Room LT: Long-Teng Conference Room at 2nd floor (二楼龙腾阁)
Room MM: Mu-Mian Meeting Room at 1st floor (一楼木棉厅)

ISB 2017 Schedule

August 20 Sunday	08:30-10:10	Highlight Session A3 (Room FH)	Session B3 (Room LT)
		Topic: Proteomics Paper IDs: 5, 77, 80, 89 Chair: Qiangfeng Zhang	Topic: Systems Biology Paper IDs: 26, 63, 64, 69, 70 Chair: Xing-Ming Zhao
	10:10-10:30	Coffee break	
	10:30-12:10	Highlight Session A4 (Room FH)	Session B4 (Room LT)
		Topic: Systems Biology Paper IDs: 9, 15, 30, 85 Chair: Jian Huang	Topic: Diseases and Drugs Paper IDs: 13, 24, 35, 38, 45 Chair: Lei Li
	12:30-13:30	Lunch	
	14:00-15:40	Highlight Session A5 (Room FH)	Session B5 (Room LT)
		Topic: Network Biology Paper IDs: 92, 58, 76, 86 Chair: Junwen Wang	Topic: Proteomics Paper IDs: 12, 31, 41, 50, 72 Chair: Zhiping Liu
	15:40-16:20	Coffee break & Poster Session	
	16:20-18:00	Highlight Session A6 (Room FH)	Session B6 (Room LT)
		Topic: Systems Biology Paper IDs: 84, 2, 1, 43 Chair: Shihua Zhang	Topic: Systems Biology Paper IDs: 19, 20, 47, 59, 65 Chair: Guanyu Wang
	18:30-20:00	Dinner	

* Subjects to revision based on further information

ISB 2017 Program

August 18-20, Shenzhen, China

August 18 (Friday) Registration

13:00-18:00 Registration, Participants arrival in Shenzhen, check in hotel, and registration package pick up (Hotel Lobby 宾馆大堂).

18:00-20:00 Reception

20:00-21:30 Board member Meeting of ORSC-CSB (Mu-Mian Meeting Room at 1st floor 一楼木棉厅)

August 19 (Saturday) Technical sessions

08:30-11:00 Registration for late arrivals (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)

08:30-08:40 Opening Session (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)
   Chair: Luonan Chen

08:40-10:10 Plenary Session P1 (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)
   Chair: Luonan Chen
08:40-09:10 *Complex Systems Analysis on Biological Systems and its Possible Applications*
   Kazuyuki Aihara
   The University of Tokyo, Japan

09:10-09:40 *Chemical Modification: A Soft Power for Biological Complex Systems*
Jiarui Wu
Shanghai Institutes for Biological Sciences, CAS, China
09:40-10:10 *Intelligent Precision Health*
Edwin Wang
University of Calgary, Canada

10:10-10:40 Coffee break

10:40-12:10 Plenary Session P2 (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)
Chair: Jiarui Wu

10:40-11:10 *Stochastic properties of fixing genes*
David Waxman
Fudan University, China
11:10-11:40 *Temporal Transcriptomic and Proteomic Landscapes of Deteriorating Pancreatic Islets in Type 2 Diabetic Rats*
Tao Xu
Institute of Biophysics, CAS, China
11:40-12:10 *PTEN Family Members in Cancer and Beyond*
Yuxin Yin
Peking University Health Science Center, China

12:30-13:30 Lunch

13:00-14:00 Board member Meeting of CSBMB Systems Biology Division (Mu-Mian Meeting Room at 1st floor 一楼木棉厅)

14:00-15:40 Highlight Session A1 (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)
Topic: Bioinformatics
Chair: Yu Xue

14:00-14:25 *VCNet: vector-based gene co-expression network construction and its application to RNA-seq data*
Zengmiao Wang[1], Huaying Fang[1], Nelson Leung-Sang Tang[2] and Minghua Deng[1]
[1]Peking University, China
[2]The Chinese University of Hong Kong, Hong Kong
Paper ID: 48
14:25-14:50 *Finding correlated pattern via high order matching for multiple sourced*

biological data

    Xi Yang, Guoqiang Han and Hongmin Cai

    South China University of Technology

    Paper ID: 54

14:50-15:15 *Critical behaviors of structural fluctuations in the native states of proteins*

    Wei Wang

    Nanjing University, China

    Paper ID: 93

15:15-15:40 *Tumor evolution and precision cancer medicine*

    Jiguang Wang

    Hong Kong University of Science and Technology, Hong Kong

    Paper ID: 87

14:00-15:40 Session B1 (Long-Teng Conference Room at 2nd floor 二楼龙腾阁)

    Topic: Network Biology

    Chair: Bairong Shen

14:00-14:20 *HISP: A Hybrid Intelligent Approach for Identifying Directed Signaling Pathways*

    Xing-Ming Zhao[1] and Shan Li[2]

    [1]Tongji University, China

    [2]Shanghai University, China

    Paper ID: 3

14:20-14:40 *KF-finder: Identification of key factors from host-microbial networks in cervical cancer*

    Jialu Hu, Yiqun Gao and Xuequn Shang

    Northwestern Polytechnical University, China

    Paper ID: 11

14:40-15:00 *Simultaneous inference of phenotype-associated genes and relevant tissues from GWAS data via Bayesian integration of multiple tissue-specific gene networks*

    Mengmeng Wu[1], Zhixiang Lin[2], Shining Ma[2], Ting Chen[1], Rui Jiang[1] and Wing Wong[2]

    [1]Tsinghua University, China

    [2]Stanford University, USA

    Paper ID: 25

15:00-15:20 *Transcriptome and Metabolic Flux Analysis Reveal Shift of Metabolic Patterns during Rice Grain Development*

    Fangzhou Shen[1], Xuting Wu[2], Luoxi Shi[1], Yangmin Chen[1], Zhuo Wang[1], Xuan Li[2]

    [1]Shanghai Jiao Tong University, China

    [2]Institute of Plant Physiology and Ecology, CAS, China

    Paper ID: 68

15:20-15:40 *Quantifying direct dependencies in biological networks by multiscale association analysis*

Jifan Shi[1], Luonan Chen[2], Tiejun Li[1], Juan Zhao[2] and Xiaoping Liu[2]
[1]Peking University, China
[2]Shanghai Institutes for Biological Sciences, CAS, China
Paper ID: 82

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

16:20-18:00 Highlight Session A2 (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)

Topic: Systems Biology
Chair: Tianshou Zhou

16:20-16:45 *Edge-based Integration of Multi-omics Data to Predict Disease State*

Rong Zeng

Shanghai Institute of Biochemistry and Cell Biology, CAS, China

Paper ID: 91

16:45-17:10 *Rethink the coding capacity of non-coding RNA: extensive translation of circRNA driven by RNA Methylation*

Zefeng Wang

Partner Institute for Computational Biology, China

Paper ID: 90

17:10-17:35 *Accurate detection of time delays and directional interactions based on time series from complex dynamical systems*

Huanfei Ma[1] and Wei Lin[2]

[1]Soochow University, China

[2]Fudan University, China

Paper ID: 46

17:35-18:00 *A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation*

Chunhe Li[1], Tian Hong[2], Chiu-Ho Webb[2], Heather Karner[2], Sha Sun[2] and Qing Nie[2]

[1]Fudan University, China

[2]University of California, Irvine, USA

Paper ID: 79

16:20-18:25 Session B2 (Long-Teng Conference Room at 2nd floor 二楼龙腾阁)

Topic: Bioinformatics
Chair: Jiguang Wang

16:20-16:40 *HetRCNA: a novel method to identify recurrent copy number alternations from heterogeneous tumor samples based on matrix decomposition framework*

Jianing Xi, Ao Li and Minghui Wang

University of Science and Technology of China, China

Paper ID: 6

16:40-17:00 *Functional Enrichment Analysis based on Long-noncoding RNA Associations*
Tun-Wen Pai and Kuo-Sheng Hung
National Taiwan Ocean University, Taiwan
Paper ID: 53

17:00-17:20 *A novel method for identifying potential disease-related miRNAs via disease-miRNA-target heterogeneous network*
Liang Ding, Minghui Wang, Dongdong Sun and Ao Li
University of Science and Technology of China, China
Paper ID: 57

17:20-17:40 *Block Spectral Clustering And Domains Association Using Signed Graphs*
Liu Ye[1], Michael Ng[1] and Stephen Wu[2]
[1]Hong Kong Baptist University, Hong Kong
[2]The Institute of Statistical Mathematics, Japan
Paper ID: 61

17:40-18:00 *Hierarchical combinatorial deep learning architecture for pancreas segmentation for medical computed tomography cancer images*
Min Fu[1], Qiuhua Liu[1], Yaobin Ou[1], Yupei Zhao[2,3] and Xinqi Gong[1]
[1]Renmin University of China
[2]Chinese Academy of Medical Sciences
[3]Peking Union Medical College
Paper ID: 71

18:30-20:00 Banquet

August 20 (Sunday) Technical sessions

08:30-10:10 Highlight Session A3 (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)
Topic: Proteomics
Chair: Qiangfeng Zhang
08:30-08:55 *PIPI: PTM-Invariant Peptide Identification Using Coding Method*
Fengchao Yu, Ning Li and Weichuan Yu
The Hong Kong University of Science and Technology, Hong Kong
Paper ID: 5

08:55-09:20 *A novel peptide specifically binding to VEGF receptor suppresses angiogenesis in vitro and in vivo*

    Bifang He, Lin Ning and Jian Huang

    University of Electronic Science and Technology of China, China

    Paper ID: 77

09:20-09:45 *MetaMHCpan, A Meta Approach for Pan-Specific MHC Peptide Binding Prediction*

    Shanfeng Zhu

    Fudan University

    Paper ID: 80

09:45-10:10 *Identification of PTMs in regulating autophagy*

    Wankun Deng and Yu Xue

    Huazhong University of Science and Technology, China

    Paper ID: 89

08:30-10:10 Session B3 (Long-Teng Conference Room at 2nd floor 二楼龙腾阁)

    Topic: Systems Biology

    Chair: Xingming Zhao

08:30-08:50 *70ProPred: a predictor for discovering sigma70 promoters based on combining multiple features*

    Wenying He and Quan Zou

    Tianjin University, China

    Paper ID: 26

08:50-09:10 *Fixation probability of a beneficial mutation conferring decreased generation time in changing environments*

    Fangshu Cui and Bo Yuan

    Shanghai Jiao Tong University, China

    Paper ID: 63

09:10-09:30 *Inference of crosstalk effects between DNA methylation and lncRNA regulation in NSCLC*

    Binhua Tang

    Hohai University, China

    Paper ID: 64

09:30-09:50 *A Draft Genome of Myospalax baileyi and its Annotation*

    Liang Li[1], Zhanyu Wang[1], Wei Feng[1], Peng Shi[2] and Lei Li[1]

    [1]Academy of Mathematics and Systems Science, CAS, China

    [2]Kunming Institute of Zoology, CAS, China

    Paper ID: 69

09:50-10:10 *Large-scale determination and characterization of cell type-specific regulatory elements in the human genome*

    Can Wang and Shihua Zhang

    Academy of Mathematics and Systems Science, CAS, China

    Paper ID: 70

10:30-10:50 Coffee break

10:30-12:10 Highlight Session A4 (**Feng-Huang Conference Room at 2nd floor 二楼凤凰阁**)

   Topic: Systems Biology

   Chair: Jian Huang

10:30-10:55 *Energy efficiency trade-offs drive nucleotide usage in transcribed regions*

   Weihua Chen[1], Guanting Lu[2], Peer Bork[3], Songnian Hu[4] and Martin Lercher[5]

   [1]Huazhong University of Science and Technology, China

   [2]The fourth millitary medical university, China

   [3]European Molecular Biology Lab, Germany

   [4]Beijing Institute of Genomics, CAS, China

   [5]University of Duesseldorf, Germany

   Paper ID: 9

10:55-11:20 *Trait and environment interplay studies through multi-scale phenotyping*

   Le Lu and JD Liu

   Monsanto Company

   Paper ID: 15

11:20-11:45 *What is the probability of replicating a statistically significant association in genome-wide association studies?*

   Wei Jiang[1], Jing-Hao Xue[2] and Weichuan Yu[1]

   [1]The Hong Kong University of Science and Technology, Hong Kong

   [2]University College London, United Kingdom

   Paper ID: 30

11:45-12:10 *Single-cell transcriptome analysis reveals widespread monoallelic gene expression in individual rice mesophyll cells*

   Wenfeng Qian

   Institute of Genetics and Developmental Biology, CAS, China

   Paper ID: 85

10:30-12:10 Session B4 (**Long-Teng Conference Room at 2nd floor 二楼龙腾阁**)

   Topic: Diseases and Drugs

   Chair: Lei Li

10:30-10:50 *A Parsimonious Model for Drug Side-effects Prediction*

   Hao Jiang[1], Yushan Qiu[2], Wai-Ki Ching[3], Wenpin Hou[3], Xiaoqing Cheng[4] and Man Yi Yim[5]

   [1]Renmin University of China, China

   [2]Shenzhen University, China

   [3]The University of Hong Kong, Hong Kong

   [4]Xi'an Jiaotong University, China

[5]New York University Shanghai, China
Paper ID: 13

10:50-11:10 *Differential networking meta-analysis of gastric cancer across Asian and American racial groups*
Wentao Dai[1], Quanxue Li[1], Bing-Ya Liu[2], Yi-Xue Li[1] and Yuan-Yuan Li[1]
[1]Shanghai Center for Bioinformation Technology, China
[2]Shanghai Jiao Tong University, China
Paper ID: 24

11:10-11:30 *Improved Flower Pollination Algorithm for Identifying Essential Proteins*
Xiujuan Lei, Ming Fang and Ling Guo
Shaanxi Normal University, China
Paper ID: 35

11:30-11:50 *Copy Number Variation Related Disease Genes*
Chocho Phd
Northwestern Polytechnical University, China
Paper ID: 38

11:50-12:10 *Multi-target drug repositioning by bipartite block-wise sparse multi-task learning*
Limin Li
Xi'an Jiaotong University, China
Paper ID: 45

12:30-13:30 Lunch

14:00-15:40 Highlight Session A5 (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)
  Topic: Network Biology
  Chair: Junwen Wang

14:00-14:25 *Epitranscriptomic RNA Modifications: Regulations and Mechanisms*
Yungui Yang
Beijing Institute of Genomics, CAS, China
Paper ID: 92

14:25-14:50 *Coarse-grained core endogenous network predicts major pancreatic developmental lineages*
Junqiang Wang[1], Ruoshi Yuan[1], Xiaomei Zhu[2] and Ping Ao[1,2]
[1]Shanghai Jiao Tong University, China
[2]Shanghai University, China
Paper ID: 58

14:50-15:15 *m6A-Driver: Identifying Context-Specific mRNA m6A Methylation-Driven Gene Interaction Networks*
Song-Yao Zhang[1], Shao-Wu Zhang[1], Lian Liu[1], Jia Meng[2] and Yufei Huang[3]
[1]Northwestern Polytechnical University, China

[2]Xi'an Jiaotong-Liverpool University, China
[3]University of Texas at San Antonio, USA
Paper ID: 76
15:15-15:40 *Interrogating RNA interactomes*
Qiangfeng Zhang
Tsinghua University, China
Paper ID: 86

14:00-15:40 Session B5 (Long-Teng Conference Room at 2nd floor 二楼龙腾阁)

Topic: Proteomics
Chair: Zhiping Liu

14:00-14:20 *Improving Conditional Random Field Model for Prediction of Protein-RNA Residue-base Contacts*
Morihiro Hayashida[1], Noriyuki Okada[1], Mayumi Kamada[2] and Hitoshi Koyano[3]
[1]National Institute of Technology, Matsue College, Japan
[2]Kyoto University, Japan
[3]Quantitative Biology Center, Riken, Japan
Paper ID: 12
14:20-14:40 *Detecting Complexes from Edge-Weighted PPI Networks via Genes Expression Analysis*
Fei Guo, Jijun Tang and Zehua Zhang
Tianjin University, China
Paper ID: 31
14:40-15:00 *Geometric and amino acid type determinants for protein-protein interaction interfaces*
Yongxiao Yang and Xinqi Gong
Renmin University of China, China
Paper ID: 41
15:00-15:20 *On the Statistical Signicance of Protein Complex*
Youfu Su, Can Zhao, Zheng Chen, Bo Tian and Zengyou He
Dalian University of Technology, China
Paper ID: 50
15:20-15:40 *PPI network analyses of human WD40 protein family systematically reveal their tendency to assemble complexes and facilitate the complex predictions*
Xu-Dong Zou, Ke An, Yun-Dong Wu and Zhi-Qiang Ye
Peking University Shenzhen Graduate School, China
Paper ID: 72

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

16:20-18:00 Highlight Session A6 (Feng-Huang Conference Room at 2nd floor 二楼凤凰阁)

   Topic: Systems Biology
   Chair: Shihua Zhang

16:20-16:45 *Proteomics toolbox for profiling cancer signaling - Toward precision medicine*

   Ruijun Tian
   Southern University of Science and Technology, China
   Paper ID: 84

16:45-17:10 *cepip: context-dependent epigenomic weighting for prioritization of regulatory variants and disease-associated genes*

   Junwen Wang
   Mayo Clinic, USA
   Paper ID: 2

17:10-17:35 *Nonequilibrium stochastic dynamics at single cell level*

   Hao Ge[1], Hong Qian[2] and Xiaoliang Sunney Xie[3]
   [1]Peking University, China
   [2]University of Washington, USA
   [3]Harvard University, USA
   Paper ID: 1

17:35-18:00 *The tipping point before pulmonary metastasis of hepatocellular carcinoma indicated by dynamic network biomarker*

   Biwei Yang[1], Meiyi Li[1], Wenqing Tang[1], Weixin Liu[2], Si Zhang[1], Luonan Chen[2] and Jinglin Xia[1]
   [1]Fudan University, China
   [2]Shanghai Institutes for Biological Sciences, CAS, China
   Paper ID: 43

16:20-18:00 Session B6 (Long-Teng Conference Room at 2nd floor 二楼龙腾阁)

   Topic: Systems Biology
   Chair: Guanyu Wang

16:20-16:40 *Modeling Biological Systems with Uncertain Kinetic Data Using Fuzzy Continuous Petri Nets*

   Fei Liu[1], Siyuan Chen[2] and Monika Heiner[3]
   [1]South China University of Technology, China
   [2]Harbin institute of technology, China
   [3]Brandenburg Technical University Cottbus-Senftenberg, Germany
   Paper ID: 19

16:40-17:00 *Detection of Driver Pathways with Rarely Mutated Genes in Cancers*

   Li Feng, Gao Lin and Wang Bingbo
   Xidian University, China
   Paper ID: 20

17:20-17:40 *Identification of novel drug targets for Diamond Black-fan Anemia (DBA)*

based on RPS19 gene mutation, using protein-protein interaction network

Abbas Khan[1], Arif Ali[1], Muhammad Junaid[1], William Cs Cho[2] and Dong-Qing Wei[1]

[1]Shanghai Jiao Tong University, China
[2]Queen Elizabeth Hospital, Hong Kong
Paper ID: 47

17:40-18:00 *A branch point on differentiation trajectory is the bifurcating event revealed by dynamic network biomarker analysis of single-cell data*

Ziwei Chen[1], Xiangqi Bai[1], Liang Ma[2], Xiawei Wang[3], Xiuqin Liu[4], Yuting Liu[5], Luonan Chen[6] and Lin Wan[1]

[1]Academy of Mathematics and Systems Science, CAS, China
[2]Beijing Institute of Genomics, CAS, China
[3]University of Wisconsin-Madison, USA
[4]University of Science and Technology Beijing, China
[5]Beijing Jiaotong University, China
[6]Shanghai Institutes for Biological Sciences, CAS, China
Paper ID: 59

18:00-18:20 *Feedback regulation in a stem cell model with acute myeloid leukaemia*

Jianfeng Jiao, Min Luo and Ruiqi Wang
Shanhai University, China
Paper ID: 65

18:30-20:00 Dinner

Poster Session

Graphical Transformation and Similarity Clustering Research for Protein Sequences

Yihong Pan, Runjie Li, Dong Qian, Shuhang Cheng and Ping Zhu
Jiangnan University, China
Paper ID: 7

Extended partial correlation for measuring direct associations in networks

Juan Zhao[1] and Luonan Chen[2]

[1]shanghai institutes for biological sciences, CAS, China
[2]Osaka Sangyo University, Japan
Paper ID: 14

Detecting the tipping point during multistep hepatocarcinogenesis by analysis of dynamic network biomarkers

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

Paper ID: 23

Revealing the tipping points during infant brain development for human and chimpanzee by analysis of gene expression data

Hui Tang[1], Ying Tang[2], Luonan Chen[2] and Tao Zeng[2]

[1]ShanghaiTech University, China

[2]Shanghai Institutes for Biological Sciences, CAS, China

Paper ID: 27

Application of deep learning in improving Recognition Network Biomarker of Complex Diseases

Minrui Peng and Luonan Chen

Institute of Biochemistry and Cell Biology, CAS, China

Paper ID: 32

Link-based Overlapping Community Detection in Complex Networks Using Symmetric Binary Matrix Factorization

Hongwei Liu

Beijing Wuzi University, China

Paper ID: 34

DNB Analysis Reveals Two Tipping Points of hESCs Differentiation Based on Single-cell RNA-seq Data

Lili Yan[1], Yuan Lou[1] and Luonan Chen[2]

[1]Renmin University of China, China

[2]Shanghai Institutes for Biological Sciences, CAS, China

Paper ID: 36

Network Analyses of Single-cell RNA-seq Data and its Application in Cancer Research

Hao Dai, Tao Zeng and Luonan Chen

Shanghai Institutes for Biological Sciences, CAS, China

Paper ID: 37

Characterizing the dynamical process and its tipping point during C3 and C4 photosynthesis by sequential transcriptome analysis

Zhonglin Jiang, Lina Lu, Luonan Chen and Tao Zeng

Shanghai Institutes for Biological Sciences, CAS, China

Paper ID: 40

Discovering dynamical network biomarkers during the onset and progression of atherosclerosis by systems biology

Jing Ge[1], Xiaoya Fan[1], Luonan Chen[1], Xinli Xue[1], Gaopeng Li[1], Shanshan Zhong[1], Xia Shen[2] and Huiyong Yin[1]

[1]Shanghai Institutes for Biological Sciences, CAS, China

[2]ShanghaiTech University, China

Paper ID: 42

Efficiently recognizing edge-biomarkers for disease sub-typing associated with survival time

Xin Zhou[1], Luonan Chen[1], Tao Zeng[1] and Yihe Yang[2]

[1]Shanghai Institutes for Biological Sciences, CAS, China

[2]Sichuan University, China

Paper ID: 44

*The above program subjects to revision based on further information.

Book of Abstracts

Plenary Sessions

Complex Systems Analysis on Biological Systems and its Possible Applications

Kazuyuki Aihara
The University of Tokyo, Japan

Biological systems are typical examples of complex systems with nonlinear spatio-temporal dynamics. In this keynote talk, first I introduce our theoretical platform for complex systems analysis [1]. Then, I explain our recent works especially on biological complex systems [2] such as hierarchical circadian dynamics in Arabidopsis [3], dynamical network markers [4], and echolocating behavior of bats [5], as well as possible engineering applications such as coherent Ising machines [6,7].

References

[1] K. Aihara, J. Imura, and T. Ueta (Eds.): "Analysis and Control of Complex Dynamical Systems: Robust Bifurcation, Dynamic Attractors, and Network Complexity," Springer, Japan (2015).

[2] L. Chen, R. Wang, C. Li, and K. Aihara: "Modeling Biomolecular Networks in Cells: Structures and Dynamics," Springer, London (2010).

[3] N. Takahashi, Y. Hirata, K. Aihara, and P. Mas: "A Hierarchical Multi-oscillator Network Orchestrates the Arabidopsis Circadian System," Cell, Vol.163, No.1, pp.148-159 (2015).

[4] R. Liu, P. Chen, K. Aihara, and L. Chen: "Identifying Early-warning Signals of Critical Transitions with Strong Noise by Dynamical Network Markers," Scientific Reports, Vol.5, Article No.17501, pp.1-13 (2015).

[5] E. Fujioka, I. Aihara, M. Sumiya, K. Aihara, and S. Hiryu: "Echolocating Bats Use Futuretarget Information for Optimal Foraging," Proceedings of the National Academy of Sciences of the United States of America, Vol.113, No.17, pp.4848-4852 (2016).

[6] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P.L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue: "A Coherent Ising Machine for 2000-node Optimization Problems," Science, Vol.354, No.6312, pp.603-606 (2016).

[7] P.L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R.L. Byer, M.M. Fejer, H. Mabuchi, and Y. Yamamoto: "A Fully-programmable 100-spin Coherent Ising Machine with All-to-all Connections," Science, Vol.354, No.6312, pp.614-617 (2016).

Name:	**Kazuyuki Aihara**	
Job:	Professor: Institute of Industrial Science, Graduate School of Information Science and Technology, and Graduate School of Engineering, The University of Tokyo (UTokyo) Director: Collaborative Research Center for Innovative Mathematical Modelling, IIS, UTokyo	
Address:	Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan E-mail: aihara@sat.t.u-tokyo.ac.jp URL: http://www.sat.t.u-tokyo.ac.jp/index.html	
Academic degree:	Ph. D of Electronic Engineering, University of Tokyo, Japan (1982)	
Professional Experience:	1982-1983 JSPS Postdoctoral Research Fellow, UTokyo, Japan 1983-1993 Assistant Professor, Lecturer, and Associate Professor, Tokyo Denki University, Japan 1993-1998 Associate Professor, UTokyo, Japan 1998-present Professor, UTokyo, Japan 2003-2009 Project Director, Aihara Complexity Modelling Project, ERATO, JST, Japan 2010-2014 Core Researcher, Aihara Innovative Mathematical Modelling Project, Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program), Japan	
Recognition:	Best Book Award for Technology and Science, Nikkan Kogyo Shinbun (1991) AI Scientific Research Award, The International Foundation for Artificial Intelligence (1992) Gold Medal Award, Tokyo Techno Forum 21 (2000) Fellow of IEICE (2006) Honorary Professor, Shanghai University (2008) Daiwa Adrian Prize (2010) JSIAM Achievement Award, JSIAM (2017).	
Current Research:	My current research interests include mathematical modeling of complex systems, fusion of dynamical systems theory and control theory, parallel distributed processing with chaotic and quantum neural networks, and time series analysis of complicated big data.	

Major Publications:

1) "A Coherent Ising Machine for 2000-node Optimization Problems," Science, Vol.354, No.6312, pp.603-606 (2016).

2) "A Fully-programmable 100-spin Coherent Ising Machine with All-to-all Connections," Science, Vol.354, No.6312, pp.614-617 (2016).

3) "Echolocating Bats Use Future-target Information for Optimal Foraging," Proceedings of the National Academy of Sciences of the United States of America, Vol.113, No.17, pp.4848-4852 (2016).

4) "A Hierarchical Multi-oscillator Network Orchestrates the Arabidopsis Circadian System," Cell, Vol.163, No.1, pp.148-159 (2015).

5) "Fisher Information for Spike-Based Population Decoding," Physical Review Letters, Vol.97, No.9, 098102 (2006).

6) "Synchronization of Eukaryotic Cells by Periodic Forcing," Physical Review Letters, Vol.96, No.14, 148102-1-4 (2006).

7) "Generalized Bienenstock-Cooper-Munro Rule for Spiking Neurons that Maximizes Information Transmission," Proceedings of the National Academy of Sciences of the United States of America, Vol. 102, pp. 5239-5244 (2005).

8) "Bridging Rate Coding and Temporal Spike Coding by Effect of Noise", Physical Review Letters, Vol. 88, No. 24, 248101 (2002).

9) "Combination of Chaotic Neurodynamics with the 2-opt Algorithm to Solve Traveling Salesman Problems", Physical Review Letters, Vol. 79, No. 12, pp. 2344-2347 (1997).

10) "Chaotic Simulated Annealing by a Neural Network Model with Transient Chaos", Neural Networks, Vol. 8, No. 6, pp. 915-930 (1995).

11) "Chaotic Neural Networks," Physics Letters A, Vol.144, No.6/7, pp.333-340 (1990).

Chemical Modification: A Soft Power for Biological Complex Systems

Jiarui Wu

Shanghai Institutes for Biological Sciences, CAS, China

Based on "Central Dogma" of molecular biology since 1960s, researchers believe that genes and proteins are basic elements for all the living things, and all the biological activities rely on these elements. However, the "Human Genome Project" in 21st century reveals that the organisms are complex systems, in which all the molecular elements interact each other to form complex molecular network. Importantly, more and more data uncover that the chemical modifications such as phosphorylation and methylation on genome and proteins play important roles for such molecular networks. These chemical modifications provide the "soft" connections to the biological elements and also provide the "soft" regulations for the functions of genes and proteins. Therefore, the study on chemical modifications of biological molecules will be a new trend for life science.

Brief CV

Dr. Jia-Rui Wu received a doctor degree from Swiss Federal Institute of Technology in Zurich, Switerland, in 1994. He was a postdoctoral fellow in Health Science Center of State University of New York from 1994-1997. Since then, he become a professor in Shanghai Institute of Biochemistry in Chinese Academy of Sciences in Shanghai. Since 2002, Dr. Wu put more efforts to promote the development of Systems Biology in China. His lab has focus on studying molecular mechanisms of type 2 diabetes and tumor with systems biology approaches. So far, he has published more than 100 research papers in international scientific journal. Dr. Wu's appointments at present are as follows: Director of Key Laboratory of Systems Biology, CAS; Deputy Dean of School of Life Science and Technology, ShanghaiTech University. Dr. Wu has positions of academic service as follows: Editor-in-Chief for "Journal of Molecular Cell Biology"; Associate Editors for "BMC Systems Biology", "Frontiers in Systems Physiology"; Members of Editorial Board for "Cell Research" and "IUBMB Life".

Intelligent Precision Health

Edwin Wang

University of Calgary, Canada

Cancer is the leading cause of death and the third largest burden in the healthcare system in the world. Each year, more than 15 million new cancer patients are diagnosed and 7-8 million people die from cancer in the world. Current precision oncology is focusing on cancer treatment, however, with some notable exceptions, improvements in overall survival and morbidity over the past few decades have been modest. Historical data suggest that early detection of cancer is crucial for its ultimate control and prevention. To meet the challenges of the surge in cancer cases in the future, it is envisioned that, besides the promotion of lifestyle changes, improving early diagnosis is the best strategy for reducing the impact of carcinogenesis.

Both genetic and environmental factors (e.g., pollution, lifestyle and so on) interact to induce cancer initiation, progression and metastasis. Therefore, we are aiming to combine the genome sequencing, imaging and electronic medical records of individuals to identify high-risk cancer individuals, 'healthy lifestyle patterns' for cancer prevention, and monitor high-risk cancer individuals for cancer early detection. To do so, we have complied a cohort which conations 5 million people whose medical records have been collected. Among them, 0.5 million people' genomic information has been determined. We are developing new algorithms by applying machine learning and deep learning approaches to the cohort to meet the goals mentioned above.

Brief CV

Edwin Wang, Professor and AISH Chair in Genomics at University of Calgary. Before that he was a professor at National Research Council Canada and McGill University. His expertise is genomics, bioinformatics and systems biology. He is an Editor for several international journals including *PLoS Computational Biology*, the top journal in the field of bioinformatics. He has edited the book of *Cancer Systems Biology* (2010), the first book of the field. He is the member of the *AACR-Cancer Systems Biology Think Tank*, which consists of ~30 world leaders in the field for discussing key problems and cutting-edge directions. His pioneering work of cancer network motifs has been featured in the college textbook, *GENETICS (2014/2017)* written by a Nobel Laureate, Dr. Hartwell and the father of systems biology, Dr. Hood. His pioneering work of microRNA of signaling networks *opens the new research area: network biology of non-coding RNAs*. He has proposed 'Cancer Hallmark Network Framework' in which cancer hallmarks can be represented, quantified and further modeled computationally using cancer hallmark networks in an evolutionary context. This framework has bene used to develop new algorithms for predicting clinical outcomes using genome sequencing data alone or integrating omic, EMR (electronic medical record) and imaging data.

Stochastic properties of fixing genes

David Waxman
Fudan University, China

In this talk I consider properties of fixation of an allele in a population. Fixation is key to understanding the way long-term evolutionary change occurs at the gene and molecular levels. It involves the frequency (or proportion) of an allele in a population achieving a value of unity. I present exact results for the way fixation occurs within the Wright-Fisher model. I give the form of a stochastic difference equation for some exactly soluble schemes of selection. The stochastic difference equation generates frequency trajectories which ultimately achieve fixation. The stochastic difference equation leads to alternative interpretation of fixation. Namely, that fixation is equivalent to a population where there is injection, each generation, of one carrier of the ultimately fixing allele into the population; repeated injection of a single allele of a given type, into a population, guarantees fixation of the allele.

Brief CV

David Waxman started his studies in physics, gaining a PhD in theoretical physics. A long-term interest in biology led to him moving from physics to biology, and in 1998 he joined the Department of Biology at Sussex University. In 2011 he joined the Centre for Computational systems Biology at Fudan University. His main interests are in evolution and population genetics, with particular emphasis on dynamics and the underlying stochastic aspects of these subjects.

Recent papers include:

- The diffusion equation of random genetic drift – biology's analogue of the Schrödinger equation?
  Contemporary Physics 2017
- Exact results for the probability and stochastic dynamics of fixation in the Wright-Fisher model
  Journal of Theoretical Biology 2017
  by Hassan Shafiey and David Waxman
- Evolutionary control: Targeted change of allele frequencies in natural populations using externally directed evolution
  Journal of Theoretical Biology 2017
  by Hassan Shafiey, Toni Gossmann and David Waxman
- The Characteristic Trajectory of a Fixing Allele: A Consequence of Fictitious Selection That Arises from Conditioning
  Genetics 2013
  by Lei Zhao, Martin Lascoux, Andrew D J Overall and David Waxman

Temporal Transcriptomic and Proteomic Landscapes of Deteriorating Pancreatic Islets in Type 2 Diabetic Rats

Tao Xu

Institute of Biophysics, CAS, China

Progressive reduction in beta-cell mass and function comprise the core of the pathogenesis mechanism of type 2 diabetes (T2D). The process of deteriorating pancreatic islets, in which a complex network of molecular events was involved, is not yet fully characterized. Here we employed RNA Sequencing and tandem mass tag (TMT) based quantitative proteomics technology to measure the temporal mRNA and protein expression changes of pancreatic islets in Goto-Kakizaki (GK) rats from 4 to 24 weeks of age. Our omics dataset outlined the dynamics of molecular network during the deterioration of GK islets as two stages: the early stage (4-6 weeks) is characterized by anaerobic glycolysis, inflammation priming, and compensation for insulin synthesis, whereas the late stage (8-24 weeks) is characterized by inflammation amplification and compensation failure. Further time-course analysis allowed us to reveal 5551 differentially expressed genes, a large portion of which have not been reported before. Our comprehensive and temporal transcriptome and proteome data offer a valuable resource for the diabetes research community and quantitative biology as well.

CURRICULUM VITAE

PRESENT INFORMATION

Name: Tao Xu, PhD

Title: Professor, Director

Institution: Institute of Biophysics

        Chinese Academy of Sciences

        Beijing 100101, P.R. China

Phone: +86-10-64888469

Fax: +86-10-64871293

E-mail: xutao@ibp.ac.cn

RESEARCH INTERESTS

Mainly focus on identifying molecular mechanisms of exocytosis and related signal transduction pathways in neuroendocrine cells.

EDUCATION

1992-1996   Ph.D. (Biophysics), Institute of Biophysics & Biochemistry, Huazhong University of Science and Technology, China.

1988-1992   B.S. (Engineering), Huazhong University of Science and Technology, China.

RESEARCH EXPERIECES

2007-   Professor, Director, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China

2003-2007   Professor, Deputy Director, Institute of Biophysics,
Chinese Academy of Sciences, Beijing, China
2000-2003   Professor, Institute of Biophysics and Biochemistry, Huazhong University of Science and Technology, China.
1999-2000   Senior Fellow, Department of Physiology and Biophysics, University of Washington, USA.
1996-1999   Postdoctoral Fellowship, Max-Planck-Institute for Biophysical Chemistry, Germany.

AWARDS
2012   The Science and Technology Awards of the Ho Leung Ho Lee Foundation
2008   National Natural Science Award (the Second Class), China
2007   Young Affiliate, TWAS

MEMBERSHIP IN ACADEMIES AND SOCIETIES
2015   Biophysics Reports, Chief Editor
2008-   Editorial Board Vice Chair, Biochemical Journal

PTEN Family Members in Cancer and Beyond

Yuxin Yin

Peking University Health Science Center, China

PTEN is essential for the maintenance of chromosomal stability and inhibition of cancer. PTEN contains an N-terminal phosphatase domain that dephosphorylates PIP3 in the cytoplasm and a C-terminal region that is important for tumor suppression. The PTEN mutant mice lacking the C-terminal domain undergo genomic instability and develop spontaneous tumors. We also found that PTEN C terminal disruption induces p53 and its downstream targets. Simultaneous depletion of p53 facilitates malignancy and promotes metastasis, suggesting that PTEN and p53 play different roles in suppression of tumor development. We have recently found that PTEN controls DNA replication progression through MCM2.   PTEN also stabilizes replication forks through RPA1.   We propose that PTEN is a regulator of DNA replication and protector of replication forks.   Our data highlights a new mechanism by which PTEN maintains genomic stability and suppresses tumorigenesis.

In our recent studies, we have revealed a mechanism of alternative protein translation, through which we identified two members of the PTEN family with novel functions. A CUG codon upstream of and in-frame with the coding region of canonical PTEN initiates translation of an N-terminally extended form of PTEN, which we have designated PTENα. We found that eukaryotic translation initiation factor 2A (eIF2A) controls PTENα synthesis and a CUG-centered palindromic motif is required in this process. PTENα induces cytochrome c oxidase activity and ATP production in mitochondria. Deletion of PTENα impairs mitochondrial respiratory chain function. Recently, we have identified another N-terminal extended PTEN isoform, designated PTENβ. PTENβ translation is initiated from an AUU codon upstream of and in-frame with the AUG initiation sequence for canonical PTEN. PTENβ localizes predominantly in the nucleolus, and regulates rDNA transcription.

Our data provide insights into the mechanism by which the PTEN family is involved in multiple cellular processes.   These results suggest that mammalian cells can use alternate translation initiation mechanisms to produce isoforms of protein with distinct functions.

Yuxin Yin, M.D., Ph.D.

University Professor and Dean

School of Basic Medical Sciences

Peking University Health Science Center

Director, Institute of Systems Biomedicine, Peking University

Dr. Yuxin Yin received his Ph.D. from the University of North Carolina at Chapel Hill in 1997 and then did his postdoctoral training in Princeton University.   In 1999, he became faculty in Columbia University Medical Center and returned to China in 2008.   His research field is cancer research with focus on the roles of tumor suppressors p53 and PTEN in cell cycle regulation, apoptosis and genomic stability. In 1992, he found a fundamental role for p53 in controlling the G1 checkpoint and maintaining genomic stability (*Cell, 1992*). In 1998, he reported that p53 is required for the cellular apoptotic

response to oxidative stress (*Nature, 1998*). In 2003, his team at Columbia demonstrated that PAC1 phosphatase is a direct transcriptional target of p53 in signaling apoptosis and tumor suppression, establishing a link between p53 and the MAP kinase cascade (*Nature, 2003*). They also found that RAD9 is a transcription factor and that RAD9 can regulate p21 and other genes involved in cell cycle checkpoints and embryogenesis (*PNAS, 2005*). In 2007, his group reported that PTEN plays a fundamental role in the maintenance of chromosome stability and that PTEN does so through physical interaction with centromeres (*Cell, 2007*). In Peking University, Dr. Yin has been investigating the PTEN pathway in genome stability and tumor suppression. In 2011, they characterized PTEN as a nuclear phosphatase of a transcription factor and identified CREB as a novel protein target of PTEN phosphatase (*Cancer Res., 2011*). His group revealed a novel mechanism of alternative translation of protein and discovered two new members of the PTEN protein family named PTENα and PTEN β （*Nature Communications*, *2017*). They found that PTEN C-terminal deletion causes genomic instability and tumor development *(Cell Reports, 2014)*. In 2014, his group reported that PTEN interacts with Histone H1 and controls chromatin condensation *(Cell Reports, 2014)*. In 2015, they revealed a new mechanism for DNA replication (*Cell Research, 2015*) and demonstrated that PTEN could controls the DNA replication process in response to replicative stress (*Cell Reports, 2015*). These studies may provide new insights into the mechanism by which PTEN maintains genomic stability and suppresses tumorigenesis. In the field of immunity, they found that PAC1/DUSP2 is a STAT3 phosphatase that modulates T_H17 development in inflammatory disease (*Nature Immunology*, 2016). Recently, their group is in-depth research and design of new drugs to inhibit tumor growth, or activate the immune system to kill tumor cells and promote tumor immunotherapy.

Parallel and Poster Sessions

Paper ID: 1

Nonequilibrium stochastic dynamics at single cell level

Hao Ge[1], Hong Qian[2] and Xiaoliang Sunney Xie[3]

[1]Peking University, China

[2]University of Washington, USA

[3]Harvard University, USA

Stochastic processes become more and more popular to model the mesoscopic nonequilibrium biophysical dynamics, which well fit the recent development of advanced experimental techniques at single-cell level.

Here I will take about two short stories. One is the molecular mechanism of transcriptional burst, which is uncovered by combining single-molecule in vitro experiments and stochastic models [1]. The other is a new rate formula for phenotype transition at the intermediate region of gene-state switching for single cells, which is more general and more close to the reality of living cells [2]. The new rate formula can explain a "noise enhancer" therapy for HIV reported recently, which motivated a future project of us.

REFERENCES

1.Chong, S.S., Chen, C.Y., Ge, H. and Xie, X.S.: Mechanism of transcriptional bursting in bacteria. Cell 158, 314–326 (2014).

2.Ge, H., Qian, H. and Xie, X.S.: Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. Phys. Rev. Lett. 114, 078101 (2015)

Paper ID: 2

cepip: context-dependent epigenomic weighting for prioritization of regulatory variants and disease-associated genes

Junwen Wang

Mayo Clinic, USA

It remains challenging to predict regulatory variants in particular tissues or cell types due to highly context-specific gene regulation. By connecting large-scale epigenomic profiles to expression quantitative trait loci (eQTLs) in a wide range of human tissues/cell types, we identify critical chromatin features that predict variant regulatory potential. We present cepip, a joint likelihood framework, for estimating a variant's regulatory probability in a context-dependent manner. Our method exhibits significant GWAS signal enrichment and is superior to existing cell type-specific methods. Furthermore, using phenotypically relevant epigenomes to weight the GWAS SNPs, we improve the statistical power of the gene-based association test.

Paper ID: 3

HISP: A Hybrid Intelligent Approach for Identifying Directed Signaling Pathways

Xing-Ming Zhao[1] and Shan Li[2]
[1]Tongji University, China
[2]Shanghai University, China

Signal transduction plays important roles in biological systems. Unfortunately, our knowledge about signaling pathways is far from complete. Specifically, the direction of signaling flows is less known even though the signalling molecules of some signaling pathways have been determined. In this paper, we propose a novel hybrid intelligent method, namely HISP (a Hybrid Intelligent approach for identifying directed Signaling Pathways), to determine both the topologies of signaling pathways and the direction of signaling flows within a pathway based on integer linear programming and genetic algorithm. By integrating the protein-protein interaction, gene expression and gene knockout data, our HISP approach is able to determine the optimal topologies of signaling pathways in an accurate way. Benchmark results on yeast MAPK signaling pathways demonstrate the efficiency of our proposed approach. When applied to the EGFR/ErbB signaling pathway in human hepatocytes, HISP unveils a high-resolution signaling pathway, where many signaling interactions are missed by existing computational approaches.

Paper ID: 5

PIPI: PTM-Invariant Peptide Identification Using Coding Method

Fengchao Yu, Ning Li and Weichuan Yu
The Hong Kong University of Science and Technology, Hong Kong

Identifying post-translational modifications (PTMs) from mass spectrometry (MS) data is a challenging tack due to the huge search space. Currently, there are two widely used approaches to addressing this issue. The first one limits the number of possible PTM types during database searching (Eng, McCormack, & Yates, 1994; Perkins, Pappin, Creasy, & Cottrell, 1999). The second one uses alignment based algorithms to infer possible modification patterns (Na, Bandeira, & Paek, 2012; Tsur, Tanner, Zandi, Bafna, & Pevzner, 2005). Both approaches have limitations. The first one only identifies a few types of PTMs. Chick et al., have already shown that there are a considerable number of modified peptides being missed with such an approach. The second one has a long running time and a low sensitivity. Most of them can only handle a small database with hundreds of proteins. These limitations obstruct biologist from identifying uncommon PTMs from the complex MS data. Some rare PTMs might be biological very important.
We try to pave the way of identifying unlimited PTMs by proposing a method named PIPI (Yu, Li, & Yu, 2016). It codes sequence database and experimental spectra into PTM-invariant vectors, respectively. Then, it uses these vectors to identify the peptides that match the experimental vectors. After obtaining the peptide sequences for each spectrum, it performs PTM characterization and localization to get the modification pattern. Finally, it performs final scoring and q-value estimating. Experiments show that PIPI identifies more peptide-spectrum matches than most state-of-the-art methods, including MS-Alignment (Tsur et al., 2005) and MODa (Na et al., 2012). Experiments also show that PIPI is much faster than all competitors, including MS-Alignment, ProteinProspector

(Chalkley, Baker, Medzhiradszky, & Burlingame, 2007), and MODa. PIPI makes it feasible to identify all possible PTMs from a standard MS data set within a few hours.

Paper ID: 6

HetRCNA: a novel method to identify recurrent copy number alternations from heterogeneous tumor samples based on matrix decomposition framework

Jianing Xi, Ao Li and Minghui Wang
University of Science and Technology of China, China

A common strategy to discovering cancer associated copy number aberrations (CNAs) from a cohort of cancer samples is to detect recurrent CNAs (RCNAs). Although the previous methods can successfully identify communal RCNAs that are shared publicly by nearly all tumor samples, detecting subgroup-specific RCNAs and their related subgroup samples from cancer samples with heterogeneity is still invalid for these existing approaches. In this paper, we introduce a novel integrated method called HetRCNA, which can identify statistical significant subgroup-specific RCNAs and their related subgroup samples. Based on matrix decomposition framework with weight constraint, HetRCNA can successfully measure the subgroup samples by coefficients of left vectors with weight constraint from multiple layers and subgroup-specific RCNAs by coefficients of the related right vectors and significance test. When we evaluate HetRCNA on simulated dataset and compare it with two previous methods, the results show that HetRCNA gives the best performances among the competing methods and is robust to the noise factors of the simulated data. When HetRCNA is applied on a real breast cancer dataset, our approach successfully identifies a bunch of RCNA regions and the result is highly correlated with the results of the other two investigated approaches. Notably, the genomic regions identified by HetRCNA harbor many breast cancer related genes which have been reported by previous researches.

Paper ID: 7

Graphical Transformation and Similarity Clustering Research for Protein Sequences

Yihong Pan, Runjie Li, Dong Qian, Shuhang Cheng and Ping Zhu
Jiangnan University, China

Sequence analysis is one of the major topics in the bioinformatics. Visualization of sequences is a new way to handle the vast amounts of biological data. It helps people to reveal the evolution relationships of different species. In this paper, we outline a new method of protein sequences graphical representation and similarity analysis. Firstly, according to the physical-chemical properties of amino acids, we get 3D coordinates of the original sequences and draw 3D graphics. Then we transform sequence graphic space coordinates into numerical sequences. We obtain the power spectra of original sequences by Discrete Fourier Transform(DFT) and extend short DFT power spectra to equal the longest length of the sequences compared. The Euclidean distance measure is employed in constructing phylogenetic tree to compare the similarities of the sequences. Our method is tested in different datasets and the results are compared with other software and algorithm. The result shows our approach is rational and effective. At last, in order to analyze

directly, we transform 3D graphic into 20D matrix diagram and compare the matrix diagrams of different species. We find that the composition of amino acid in various sequences is different. The new method of graphical representation and similarity clustering research is of certain significance for promoting the development of bioinformatics.

Paper ID: 9

Energy efficiency trade-offs drive nucleotide usage in transcribed regions

Weihua Chen[1], Guanting Lu[2], Peer Bork[3], Songnian Hu[4] and Martin Lercher[5]
[1]*Huazhong University of Science and Technology, China*
[2]*The fourth millitary medical university, China*
[3]*European Molecular Biology Lab, Germany*
[4]*Beijing Institute of Genomics, CAS, China*
[5]*University of Duesseldorf, Germany*

Efficient nutrient usage is a trait under universal selection. A substantial part of cellular resources is spent on making nucleotides. We thus expect preferential use of cheaper nucleotides especially in transcribed sequences, which are often amplified thousand-fold compared with genomic sequences. To test this hypothesis, we derive a mutation-selection-drift equilibrium model for nucleotide skews (strand-specific usage of 'A' versus 'T' and 'G' versus 'C'), which explains nucleotide skews across 1,550 prokaryotic genomes as a consequence of selection on efficient resource usage. Transcription-related selection generally favours the cheaper nucleotides 'U' and 'C' at synonymous sites. However, the information encoded in mRNA is further amplified through translation. Due to unexpected trade-offs in the codon table, cheaper nucleotides encode on average energetically more expensive amino acids. These trade-offs apply to both strand-specific nucleotide usage and GC content, causing a universal bias towards the more expensive nucleotides 'A' and 'G' at non-synonymous coding sites.

Paper ID: 11

KF-finder: Identification of key factors from host-microbial networks in cervical cancer

Jialu Hu, Yiqun Gao and Xuequn Shang
Northwestern Polytechnical University, China

Background: The human body is colonized by a vast number of microbes. Microbiota can benefit many normal life processes, but can also cause many diseases by interfering the regular metabolism and immune system. It is known that oncogenic human papillomavirus is one of the major causes of cervical carcinoma. However, the effect of microbiota in cervical cancer is still unclear. Therefore, it is of significance to understand the regulatory mechanism of microbiota in uterine cervix cancer. Results: In this paper, we investigated microbiota composition and gene expression data for 58 squamous and adenosquamous cell carcinoma, and reconstructed a host-microbial network, which consists of 259 abundant microbes and 738 differentially expressed genes (DEGs). A variety of meta-analysis methods were performed to search for key risk factors from the host-microbial networks. A web-based tool KF-finder was developed to efficiently query and visualize the

knowledge of microbiota and differentially expressed genes (DEGs) in cervical cancer. The results indicate that prevotellaceade, tissierellaceae and fusobacteriaceae are the most abundant microbes in cervical carcinoma, and the microbial community in cervical cancer is less diverse than that of any other boy sites. Several key risk factors anaerococcus, hydrogenophilaceae, eubacterium, PSMB10, KCNIP1 and KRT13 were found, which can regulate the progress of viral response, cell cycle and epithelial cell differentiation in cervical cancer.

Conclusions: Besides oncogenic HPVs, the interplay between microbiota and differentially expressed genes may play crucial roles in the development of cervical cancer, particularly these interactions with key risk factors. Permanent changes of microbiota composition could be a major reason for chromosomal instability, which subsequently enables the effect of key risk factors in cancer. All our results described in this paper can be freely accessed from our website at http://www.nwpu-bioinformatics.com/KF-finder/.

Paper ID: 12

Improving Conditional Random Field Model for Prediction of Protein-RNA Residue-base Contacts

Morihiro Hayashida[1], Noriyuki Okada[1], Mayumi Kamada[2] and Hitoshi Koyano[3]
[1]National Institute of Technology, Matsue College, Japan
[2]Kyoto University, Japan
[3]Quantitative Biology Center, Riken, Japan

For understanding biological cellular systems, it is important to analyze interactions between protein residues and RNA bases. A method based on conditional random fields (CRFs) was developed for predicting contacts between residues and bases, which receives multiple sequence alignments for given protein and RNA sequences, respectively, and learns the model with many parameters involved in relationships between neighboring residue-base pairs by maximizing the pseudo likelihood function.

In this paper, we propose a novel model with less parameters based on CRFs. For evaluating the proposed model, we performed cross-validation experiments, and took the average of AUC (area under receiver operating characteristic curve) scores. The result suggests that the proposed CRF-based model without using L1-norm regularization (lasso) outperforms the existing model with and without the lasso under several input observations to CRFs.

Paper ID: 13

A Parsimonious Model for Drug Side-effects Prediction

Hao Jiang[1], Yushan Qiu[2], Wai-Ki Ching[3], Wenpin Hou[3], Xiaoqing Cheng[4] and Man Yi Yim[5]
[1]Renmin University of China, China
[2]Shenzhen University, China
[3]The University of Hong Kong, Hong Kong
[4]Xi'an Jiaotong University, China
[5]New York University Shanghai, China

The identification of drug side-effects is considered to be an important step in drug design, which could not only shorten the time but also reduce the cost of drug development. In this paper, we investigate the relationship between the potential side-effects of drug candidates and their chemical structures. First, we propose a Naive Regression (NR) model for drug side-effects prediction.

Its promising features lie in the efficiency of training model parameters and the existence of a closed for solution.

Meanwhile, the primary foundation of regression creates a golden opportunity for evaluating drug side-effect profiles efficiently.

Then an improved version, Regularized Regression (RR) model, further increases the prediction accuracy.

The usefulness of NR and RR is demonstrated in a cross validation setting on a 1358 side-effects profiling from SIDER database and the chemical structures of 888 approved drugs.

The model RR, as anticipated, remarkably outperforms NR which at the same time performs better than other state-of-the-art methods (decision tree, k-nearest neighbors algorithm, and support vector machines), in efficiency as well as accuracy.

Paper ID: 14

Extended partial correlation for measuring direct associations in networks

Juan Zhao[1] and Luonan Chen[2]
[1]*shanghai institutes for biological sciences, CAS, China*
[2]*Osaka Sangyo University, Japan*

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: 15

Trait and environment interplay studies through multi-scale phenotyping

Le Lu and JD Liu
Monsanto Company

Many physiological and molecular traits play an important role in crop yield. Trait performance is impacted by both environmental variation and intrinsic genetic variation, which presents a key

challenge with identifying the most important traits to modify for yield improvement. We conducted a systemic experiment with collecting many field phenotypes, metabolites and transcripts on a set of 57 diverse, commercially relevant maize hybrids across 3 environments to examine the contribution of environmental and genetic factors to the variance of these traits, and to understand their relations to yield and to each other. We found that for yield and yield components, biomass, roots, and nitrogen traits environmental and residual effects represent the largest proportion of variance, while genetic factors contribute more significantly to phenology traits, plant height and kernel row number. Only a small number of genes showed expression correlating to the traits across environments, and phenology and nitrogen-content traits have the most consistent correlation with yield across years.

Paper ID: 19

Modeling Biological Systems with Uncertain Kinetic Data Using Fuzzy Continuous Petri Nets

Fei Liu[1], Siyuan Chen[2] and Monika Heiner[3]
[1]*South China University of Technology, China*
[2]*Harbin institute of technology, China*
[3]*Brandenburg Technical University Cottbus-Senftenberg, Germany*

Background: Uncertainties exist in many biological systems, which can be classified as random uncertainties and fuzzy uncertainties. The former can usually be dealt with using stochastic methods, while the latter has to be handled with such methods as fuzzy methods.

Results: In this paper, we focus on a special type of biological systems, which can be described using ordinary differential equations (ODEs) or continuous Petri nets (CPNs), but some kinetic parameters are missing or inaccurate. For this, we propose a class of fuzzy continuous Petri nets (FCPNs) by combining CPNs and fuzzy logics. We also present and implement a simulation algorithm for FCPNs, and illustrate our method with the heat shock response model.

Conclusion: This approach can be used to model a biological system where some of its kinetic parameters are not available or their values vary due to some environmental factors.

Paper ID: 20

Detection of Driver Pathways with Rarely Mutated Genes in Cancers

Li Feng, Gao Lin and Wang Bingbo
Xidian University, China

Identifying driver pathways is a key challenge to interpret the molecular mechanisms and pathogenesis underlying cancer. An increasing number of studies suggest that rarely mutated genes are important for the development of cancer. At present, most approaches have been proposed to identify driver pathways with frequently mutated genes. However, the driver pathways consisting of mutated genes with low-frequency driver mutations are not well characterized. To identify driver pathways with rarely mutated genes, we first propose a functional similarity index via quantifying the functional relationship between mutated driver genes within a pathway. Then, we develop a method to detect Driver Pathways with Rarely Mutated Genes (aka, DPRMG), which incorporates

the functional similarity, coverage and mutual exclusivity of mutations. By applying DPRMG on TCGA cancer dataset, we detect driver pathways, which intersect with the well-known signalling pathways and protein complexes. Furthermore, one obtained driver pathway is enriched with the cell cycle pathway, in which several cancer genes with low frequencies, including ESR1, HDAC1, SMAD3 and STAT3. Meanwhile we identify the mediator complex with rarely mutated genes MED1, MED15 and MED14. When compared with HotNet2 on three networks: HINT+HI2012 MultiNet and iRefIndex, DPRMG detects more rarely mutated cancer genes and has higher F-measure. Overall, DPRMG provides an effective method for the identification of driver pathways with rarely mutated genes.

Paper ID: 23

Detecting the tipping point during multistep hepatocarcinogenesis by analysis of dynamic network biomarkers

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

Hepatocellular carcinoma (HCC) is a complex disease with multi-step carcinogenic process from preneoplastic lesions, including cirrhosis, low-grade dysplastic nodules (LGDNs) and high-grade dysplastic nodules (HGDNs) to HCC. There is only an elemental understanding of its molecular pathogenesis and it is a key problem to identify when and how the critical transition happens in HCC initiation period at a molecular level. We revealed that LGDNs is the tipping point (termed pre-HCC state) of hepatocarcinogenesis based on the serial gene expression profiles by a mathematical model termed dynamic network biomarkers (DNB). Different from the conventional biomarker approach based on the abundance of molecular expressions, the DNB model exploits collective fluctuations and correlations of different genes. The large number of differentially expressed genes between cirrhosis and HGDNs was detected which highlighted the stark differences before and after tipping point, implying DNB as the early-warning signals of HCC for marking the upcoming drastic deterioration. We further identified biologic pathways responsible for this transition. Pathways related to immune system reactions, cell proliferation and apoptosis, and cell adhesion playing key role in early carcinogenesis were observed by the functional analysis of DNB. Furthermore, DNB was effective as predictor of prognostic for HCC patients, suggesting a potential clinical application of DNB.

Paper ID: 24

Differential networking meta-analysis of gastric cancer across Asian and American racial groups

Wentao Dai[1], Quanxue Li[1], Bing-Ya Liu[2], Yi-Xue Li[1] and Yuan-Yuan Li[1]
[1]Shanghai Center for Bioinformation Technology, China
[2]Shanghai Jiao Tong University, China

Gastric Carcinoma is one of the most lethal cancer around the world, and is also the most common cancers in Eastern Asia. A lot of differentially expressed genes have been detected as being

associated with Gastric Carcinoma (GC) progression, however, little is known about the underlying dysfunctional regulation mechanisms. To address this problem, we previously developed a differential networking approach that is characterized by involving differential coexpression analysis (DCEA), stage-specific gene regulatory network (GRN) modelling and differential regulation networking (DRN) analysis.

Considering the complexity and diversity of gastric carcinogenesis, we collected three datasets (GSE54129, GSE24375 and TCGA-STAD) for Chinese, Korean and American, and aimed to investigate the common dysregulation mechanisms of gastric carcinogenesis across racial groups. First, we constructed conditional GRNs for gastric cancer corresponding to normal and carcinoma, and then prioritized differentially regulated genes (DRGs) and gene links (DRLs) from three datasets separately by using our previously developed differential networking method. Based on our integrated differential regulation information from three datasets and prior knowledge (e.g., transcription factor (TF)-target regulatory relationships and known signaling pathways), we generated testable hypotheses on the regulation mechanisms of two genes, XBP1 and GIF, out of 16 common cross-racial DRGs in gastric carcinogenesis.

The current cross-racial integrative study from the viewpoint of differential regulation networking provided useful clues for understanding the common dysfunctional regulation mechanisms of gastric cancer progression and discovering new universal drug targets or biomarkers for gastric cancer.

Paper ID: 25

Simultaneous inference of phenotype-associated genes and relevant tissues from GWAS data via Bayesian integration of multiple tissue-specific gene networks

Mengmeng Wu[1], Zhixiang Lin[2], Shining Ma[2], Ting Chen[1], Rui Jiang[1] and Wing Wong[2]
[1]*Tsinghua University, China*
[2]*Stanford University, USA*

Although genome-wide association studies (GWAS) have successfully identified thousands of genomic loci associated with hundreds of complex traits in the past decade, the debate about such problems as missing heritability and weak interpretability has been appealing for effective computational methods to facilitate the advanced analysis of the vast volume of existing and anticipated genetic data. Towards this goal, gene-level modelling with the assumption that genes associated with a phenotype tend to be enriched in a gene network has recently attracted much attention, due to such advantages as easy interpretation, less multiple testing burdens, and robustness across studies. However, existing methods in this category usually exploit none tissue-specific gene networks and thus lack the ability of utilizing informative tissue-specific characteristics. To overcome this limitation, we proposed a Bayesian approach called SIGNET to integrate GWAS data and multiple tissue-specific gene regulatory networks for the simultaneous inference of phenotype associated genes and relevant tissues. With a rigorous Markov random field (MRF) model, our method automatically discovers tissues relevant to a phenotype of interest, thereby not only eliminating the need for manually selecting specific tissues but also enhancing the power of prioritizing candidate genes. Through extensive simulation studies, we showed the effectiveness of our method in finding both relevant tissues and associated genes for a phenotype. In applications to

real GWAS data of 14 complex phenotypes, we demonstrated the power of our method in both deciphering genetic basis and discovering biological insights of a phenotype. With this understanding, we expect to see SIGNET as a valuable tool for integrative GWAS analysis, thereby boosting the prevention, diagnosis and treatment of human inherited diseases and eventually facilitating precision medicine.

Paper ID: 26

70ProPred: a predictor for discovering sigma70 promoters based on combining multiple features

Wenying He and Quan Zou
Tianjin University, China

Promoter is an important sequence regulation element, which is in charge of gene transcription initiation. The promoter recognition has been a crucial part of the gene structure recognition. It's also the core issue of constructing gene transcriptional regulation network. With the successfully implement and complement of Human Genome Project, the systematic and precise identification of promoter regions in DNA sequence is a big challenge. In this study, we developed a promoter recognition method called 70ProPred by combining position-specific trinucleotide propensity based on single-stranded characteristic (PSTNPss) with the electron-ion interaction potential (EIIP) values for trinucleotides to predict sigma70 promoters in prokaryote. Finally 79 features of PSTNPss combined with 64 features of PseEIIP yielded the best performance for identifying sigma70 promoters and non-promoters, with an overall accuracy (Acc) of 95.56% and a Matthews correlation coefficient (MCC) of 0.90. Rigorous jackknife tests showed that 70ProPred was significantly better than the existing sigma70 promoter prediction methods in both overall accuracy and stability. This approach can also be extended to predict other species promoters. For the convenience of experimental biologists, an online webserver for the proposed method was built, which is freely available at http://121.42.167.206/70ProPred/index.jsp.

Paper ID: 27

Revealing the tipping points during infant brain development for human and chimpanzee by analysis of gene expression data

Hui Tang[1], Ying Tang[2], Luonan Chen[2] and Tao Zeng[2]
[1]ShanghaiTech University, China
[2]Shanghai Institutes for Biological Sciences, CAS, China

Postpartum developmental delay has been proposed as a possible characteristic of human evolution which contributes to many human-specific brain features, such as an increase in brain size and the advanced human-specific cognitive traits. However, the biological processes and molecular functions during early brain development remain poorly understood, especially in human and primates. In this work, we extensively studied gene expression data of dorsolateral prefrontal cortex in human and chimpanzee to investigate the critical biological processes or events during early brain development at a molecular level. By dynamic network biomarkers, we found that there are tipping

points around 3 months and 1 month, respectively, which are crucial periods in infant human and chimpanzee brain development. Particularly, the human postnatal development and its expression changes are delayed 3 times relative to chimpanzee, and many common biological processes are involved in those key periods for both human and chimpanzee, e.g., physiological system development functions, nervous system development, organismal development and tissue morphology. These findings support that the maximal rates of brain growth will achieve in those two critical periods for human and primates respectively. In addition, our analytic results also reveal that human starts to develop a number of advanced behavior functions around its tipping point (around 3 months), such as the ability of learning and memory.

Paper ID: 30

What is the probability of replicating a statistically significant association in genome-wide association studies?

Wei Jiang[1], Jing-Hao Xue[2] and Weichuan Yu[1]
[1]*The Hong Kong University of Science and Technology, Hong Kong*
[2]*University College London, UK*

The goal of genome-wide association studies (GWASs) is to discover genetic variants associated with diseases/traits. Replication is a common validation method in GWASs. We regard an association as true finding when it shows significance in both primary and replication studies. A question worth pondering is what is the probability of a primary association (i.e. a statistically significant association in the primary study) being validated in the replication study? This article proposes a Bayesian probabilistic measure, named the replication rate (RR), to answer this question. We further propose an estimation method for RR, which makes use of the summary statistics from the primary study. We can use the estimated RR to determine the sample size of the replication study and to check the consistency between the results of the primary study and those of the replication study. Simulation and real data experiments show that the estimated RR has good prediction and calibration performance. We also use these data to demonstrate the usefulness of RR. The R-package is available at http://bioinformatics.ust.hk/RRate.html.

Paper ID: 31

Detecting Complexes from Edge-Weighted PPI Networks via Genes Expression Analysis

Fei Guo, Jijun Tang and Zehua Zhang
Tianjin University, China

Identifying complexes from PPI networks has become a key problem to elucidate protein functions and identify signal and biological processes in a cell. Proteins are important roles of life activity, after binding as complexes. Accurate determination of complexes in PPI networks is crucial for understanding principles of cellular organization. We propose a novel method to identify complexes on PPI networks, based on different co-expression information. First, we use Markov Cluster Algorithm with an edge-weighting scheme to calculate complexes on PPI networks. Then, we propose some significant features, such as graph information and gene expression analysis, to filter

and modify complexes predicted by Markov Cluster Algorithm. To evaluate our method, we test on two experimental yeast PPI networks. On DIP network, our method has Precision and F-Measure values of 0.6004 and 0.5528. On MIPS network, our method has F-Measure and Sn values of 0.3774 and 0.3453. Comparing to existing methods, our method improves Precision value by at least 0.1752, F-Measure value by at least 0.0448, Sn value by at least 0.0771. Experiments show that our method achieves better results than some state-of-the-art methods for identifying complexes on PPI networks, with the prediction quality improved in terms of evaluation criteria.

Paper ID: 32

Application of deep learning in improving Recognition Network Biomarker of Complex Diseases

Minrui Peng and Luonan Chen
Institute of Biochemistry and Cell Biology, CAS, China

Network Biomarkers are become more and more widely used as the analysis of complex biology process goes deeper in recent years. Correlation or interactions in the network are vitally important as they are the basic structure of a network. Therefore, estimating correlation is essential, which requires more repeated experiments or larger sample in a single phase or state of the entire process. Although the revolutionary development of sequencing and omics technology have made it less expensive than before, conducting repeated experiments with large sample size can be difficult or even not likely possible due to various practical facts during the experiment process. To settle this inconsistency, we are coming up a method to predict correlation through only one sample by using deep learning framework to discover the hidden connections between genes and there interactions. Our method is based on the hypnosis that a biological system is stable, fully-measurable and partially restricted unless the system has changed. Under such circumstance, we use the microarray-based GEO datasets, consisting ~110,000 human gene expression profiles to train our ANN-based deep learning model to estimate the correlations in network. For each profile, there are ~22,000 genes in the genome, and at most ~55,000 meaningful correlations in regulatory network, which is a rather reasonable ratio in the field of machine learning. We use mean absolute error to fulfil the optimization of estimation in the training part, and to evaluate the predictive performance of each target correlation. With the estimated correlations, we are able to construct a network biomarker with only one sample at a time. Furthermore, we are conducting application to combine this research with traditional methods, seeking to improve the performance and discriminant validity, especially in the scene of time series analysis.

Paper ID: 34

Link-based Overlapping Community Detection in Complex Networks Using Symmetric Binary Matrix Factorization

Hongwei Liu
Beijing Wuzi University, China

Identification of overlapping community structures is an important topic and issue to understanding

the structure and dynamics of many networks since communities often play important roles in network systems. Communities are often defined as groups of related nodes or links that correspond to functional subunits in the corresponding complex systems. This paper proposes a symmetric binary matrix factorization model to discover overlapping communities by partitioning links. Since links usually represent unique relations among nodes, the link clustering will discover groups of links that have the same characteristics. Thus nodes naturally belong to multiple communities. Experiments on both artificial networks and real networks validate the effectiveness and efficiency of the proposed algorithm.

Paper ID: 35

Improved Flower Pollination Algorithm for Identifying Essential Proteins

Xiujuan Lei, Ming Fang and Ling Guo
Shaanxi Normal University, China

Essential proteins are necessary for the survival and development of cells. The identification of essential proteins can help to understand the minimal requirements for cellular life and it also plays an important role in the disease genes study and drug design. With the development of high-throughput techniques, a large amount of protein-protein interactions data is available to predict essential proteins at the network level. So far, even though a number of essential protein discovery methods have been proposed, the prediction precision still needs to be improved. In this paper, we propose a new algorithm, named FPE (Improved Flower Pollination Algorithm for Identifying Essential Proteins), which integrates gene expression data, subcellular localization and protein complexes information with the topological properties of PPI networks. Different from other existing methods, we use FPA which is a new nature-inspired algorithm considering the characteristics of flower pollination to discover a candidate essential protein set for the identification of essential proteins. Moreover, we develop a GSC measurement in order to judge the essentiality of proteins, which taking into account not only the Gene expression data, Subcellular localization and protein Complexes information, but also the network topological properties. The experimental results show that the proposed algorithm has high level of stability, as well as performs better than the state-of-the-art methods (DC, SC, IC, EC, LAC, NC, PeC, WDC, UDoNC and SON) in terms of the prediction precision for identifying essential proteins.

Paper ID: 36

DNB Analysis Reveals Two Tipping Points of hESCs Differentiation Based on Single-cell RNA-seq Data

Lili Yan[1], Yuan Lou[1] and Luonan Chen[2]
[1]Renmin University of China, China
[2]Shanghai Institutes for Biological Sciences, CAS, China

How to detect the early-warning signals of a biological process has attracted more and more attention from biologists. The theory of dynamic network biomarker (DNB) opens a new way to quantitatively identify the critical state or the tipping point from the observed data just before the

critical transition of the biological process. DNB can signal the imminent transition and is a dynamic network of biomarkers, which are composed of a dominant group of genes or molecules. The network pattern in DNB members is formed by the strong correlations and also fluctuations of genes during a biological process. On the other hand, recently single cell sequencing technique has been applied to analyze the cell differentiation due to the significant advantage to represent heterogeneous features of cell population. In particular, and to understand the process of cell differentiation and reveal the regulations of cell fate, it is an important to study the differentiation process at a molecular level from human pluripotent cells to definite endoderm (DE) via mesendoderm (ME). This differentiation is considered not always smooth but sometimes abrupt, thus having the tipping point. In other words, there is a critical state during this differentiation process just before their differentiation from one state to another, which can be analyzed by DNB theory from single cell sequencing data. In this work, we analyzed 758 single-cells RNA-seq data covering 6 time-points with 3 differentiation states by DNB model, which identified two tipping points as well as the respective DNB members. We also found a small group of functionally related genes and network patterns that could act as the drivers of the differentiation based on DNB members and DEG members (differential expressed genes). Furthermore, we found 5 master regulators or genes that are strongly related to the fate of the cell differentiation.

Paper ID: 37

Network Analyses of Single-cell RNA-seq Data and its Application in Cancer Research

Hao Dai, Tao Zeng and Luonan Chen
Shanghai Institutes for Biological Sciences, CAS, China

Single-cell analyses based on RNA-seq facilitate the ability to look at the changes in gene expression over time, or the differences in different cells finely. However, gene expression is regulated by many factors and the dynamic process makes the gene expression change in a large range, which sometimes causes the difficulties in expression analysis. Here we proposed a new method of single-cell analyses from the view of gene-gene interactions. Network features that represented the degree or valency of each gene in gene expression network of single cell were calculated. The method was applied in cancer research based on the data from TCGA database. Some important genes in gene expression network were found and implied some new mechanisms of cancer development.

Paper ID: 38

Copy Number Variation Related Disease Genes

Chocho Phd
Northwestern Polytechnical University, China

One of the most important and challenging issues in biomedicine and genomics is how to identify disease-related genes. Datasets from high-throughput biotechnologies have been widely used to overcome this issue from various perspectives e.g. epigenomics, genomics, transcriptomics, proteomics, metabolomics, etc. At the genomic level, copy number variations (CNVs) have been recognized as critical genetic variations, which contribute significantly to genomic diversity. They

have been associated with both common and complex diseases, and thus have a large influence on a variety of Mendelian and somatic genetic disorders.

Reliable detection of CNVs will not only allow to discriminate driver mutations for various diseases, but also helps to develop personalized medicine when integrating it with other genomic features. In this review, based on complex diseases, we give an overview about the critical role of using CNVs for identifying disease-related genes. A set of definition of CNVs and their relationship with genomic diseases are first discussed. Then, various CNV detection methods are summarized before introducing a set of CNV related genetic disease and their implicated genes and loci.

The review concludes by explaining the important role of CNVs in analyzing genetic diseases, especially when integrating them with other genomic datasets. Some limitations and challenges concerned CNVs have been summarized.

Paper ID: 40

Characterizing the dynamical process and its tipping point during C3 and C4 photosynthesis by sequential transcriptome analysis

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

C4 and C3 photosynthesis differ in the efficiency on consuming water and nitrogen, especially C4 would be a more effective way but the underlying molecular mechanism is still unclear. To take insights into the distinct molecular mechanisms involved in C4 and C3 photosynthetic construction, we carried on systematic methods to analyze the leaves development to increase crop yields related to C4 and C3 respectively. Firstly, we obtained the gene expression profiles in the developing leaves of Zea mays (a C4 plant Maize), and Oryza sativa (a C3 plant Rice), respectively by RNA-sequencing. Then, we used WGCNA to capture the expression correlation patterns among genes, and constructed a co-expression network across leaves development and selected a significant locus-related gene modules along with the leaves developing trajectories. Finally, we applied dynamic network biomarkers (DNB) to identify a critical transition stage during the leaves developing trajectories. Combined with these analysis results, we found that both maize and rice actually exist state transition during the C4 and C3 photosynthetic construction respectively; and the critical point of such transition is related with the change of sun lighting on leaves, which is also supported by the functional enrichment analysis both on gene module and DNB; and especially, the key genes selected from DNB based on the topological change of co-expression networks before and after the critical point, show significant network degree change and would be driver candidates of C4 photosynthesis with natural selection across C4 and C3 plant species.

Paper ID: 41

Geometric and amino acid type determinants for protein-protein interaction interfaces

Yongxiao Yang and Xinqi Gong
Renmin University of China, China

Motivation: Protein-protein interactions are essential to many biological processes. The binding site

information of protein-protein complexes is extremely useful to obtain their structures from biochemical experiments. Geometric description of protein structures is the precondition of protein binding site prediction and protein-protein interaction analysis. The previous description of protein surface residues is incomplete, and little attention are paid to the implication of residue types for binding site prediction.

Results: Here, we found three new geometric features to characterize protein surface residues which are very effective for protein-protein interface residue prediction. The new features and several commonly used descriptors were employed to train millions of residue type-nonspecific or specific protein binding site predictors. The amino acid type-specific predictors are superior to the models without distinction of amino acid types. The performances of the best predictors are much better than those of the sophisticated methods developed before. The results demonstrate that the geometric properties and amino acid types are very likely to determine if a protein surface residue would become an interface one when the protein binds to its partner.

Paper ID: 42

Discovering dynamical network biomarkers during the onset and progression of atherosclerosis by systems biology

Jing Ge[1], Xiaoya Fan[1], Luonan Chen[1], Xinli Xue[1], Gaopeng Li[1], Shanshan Zhong[1], Xia Shen[2] and Huiyong Yin[1]
[1]*Shanghai Institutes for Biological Sciences, CAS, China*
[2]*ShanghaiTech University, China*

Atherosclerosis is one of the major factors causing cardiovascular diseases including myocardial infarction. The onset and progression of atherosclerosis is a nonlinear dynamical process, which involves complicated dynamic regulations among bio-molecular networks. To investigate such disease onset and progression procedure, we have combined omics data and theoretical analysis to establish a theoretical method for predicting early event of the dys-regulation of blood vessel homeostasis and atherosclerotic lesion formation from the viewpoint of systems biology. Particularly, we firstly used the LDLR-/-mouse model feeding western diet to simulate the onset and progression of atherosclerosis, and the pathological characteristics of aortic roots by oil red and H&E staining showed that the atherosclerotic lesion gradually increased over time. Secondly, we integrated high-throughput data including different-stage data of RNA-seq and lipidomics to infer regulatory networks of atherosclerosis and identify dynamical network biomarkers (DNB) to characterize the critical transition from normal vessel to atherosclerotic lesion. By analyzing the RNA-seq data of aortas, we identified the critical point in the progression of atherosclerosis at the genetic level and discovered a group of DNB genes, which should play driving roles in the progression of disease. In order to make the DNB genes more easily be detected and finally be applied in clinic, we also integrated lipidomics data of plasma to discover DNB lipids at the level of small molecules. Finally, we will further validate the theoretical results with biological experiments. In summary, we aim to develop a prediction model to identify the early signal of the critical transition in atherosclerosis and look forward to provide guidance for the early prediction and prevention of atherosclerosis.

Paper ID: 43

The tipping point before pulmonary metastasis of hepatocellular carcinoma indicated by dynamic network biomarker

Biwei Yang[1], Meiyi Li[1], Wenqing Tang[1], Weixin Liu[2], Si Zhang[1], Luonan Chen[2] and Jinglin Xia[1]
[1]Fudan University, China
[2]Shanghai Institutes for Biological Sciences, CAS, China

Metastasis is one major factor for the high mortality of hepatocellular carcinoma (HCC). Thus, it is strongly demanded to find a predictive biomarker to detect the tipping point or early-warning signals before metastasis and recurrence so as to prevent the serious deterioration of the disease, which, however, is not yet available and actually a rather difficult task. Here, to discover early-warning signals of pulmonary metastasis for HCC as well as its predictive biomarkers, we analyzed the time-series gene expression data of xenograft nude mice model HCCLM3-RFP with our novel dynamical network biomarker (DNB) method. Both functional analyses and experiment validations indicate that CALML3 is one of DNB members, which indicates metastasis initiation at the critical transition stage (i.e., the 3rd week after orthotopic transplant). In particular, we found that CALML3 is a suppressor in protecting from metastasis and revealed its metastasis-related roles at a network level, which include direct or proximal regulations as well as cascading or remote influences. Importantly, as demonstrated from our biological experiments and clinical samples, the expression level of CALML3 can predict the metastasis or detect its early-warning signals, and further indicate degree of tumor malignancy for HCC, compared with the effect of typical clinical pathological factors. Thus, from clinical viewpoint, this study on the tipping point of metastasis not only provides predictive biomarkers for its early diagnosis, but also reveals the molecular mechanisms for its early prevention.

Paper ID: 44

Efficiently recognizing edge-biomarkers for disease sub-typing associated with survival time

Xin Zhou[1], Luonan Chen[1], Tao Zeng[1] and Yihe Yang[2]
[1]Shanghai Institutes for Biological Sciences, CAS, China
[2]Sichuan University, China

It has been widely accepted that genes perform their functions by working interactively with others, and these gene interactions or relationships can be abstracted as edges of a bio-molecular network. In translational bio-medical area, there were a large number of research works in finding gene biomarkers of human diseases, however, the gene-pair or edge biomarkers still require efficient detection algorithm. To address this issue, we first establish a gene-gene correlation network space, which will produce an ultrahigh dimensional feature space. We then use a sure screening method based on correlation learning to reduce the dimensionality of gene-pairs from high to a moderate scale (e.g. theory guaranteed to below the sample size), and next Smoothly clipped absolute deviation (SCAD) method is used to select important features to discriminate different phenotypes. Such whole edge-biomarker detection algorithm is called SIS-SCAD, and has been applied on five real datasets with binary phenotypes. To demonstrate the superiority of our approach over traditional

methods, we have used SIS-SCAD and six state-of-the-art methods to test the same datasets. All computational results show that our method out-performs other methods in terms of accuracy and efficiency. The identified edge-biomarkers matches well with the associated diseases, and the survival analysis shows that the selected gene-pairs rather than single genes can well distinguish the two phenotypic data. The analysis of human thyroid dataset and stomach dataset also suggested that the identified edge-biomarkers may cast new biological insights into the pathogenesis of human complex diseases.

Paper ID: 45

Multi-target drug repositioning by bipartite block-wise sparse multi-task learning

Limin Li
Xi'an Jiaotong University, China

Background
Finding potential drug targets is a crucial step in drug discovery and development. Recently, resources such as the Library of Integrated Network-Based Cellular Signatures (LINCS) L1000 database provide gene expression profiles induced by various chemical and genetic perturbations and thereby make it possible to analyze the relationship between compounds and gene targets at a genome-wide scale. Current approaches for comparing the expression profiles are based on pairwise connectivity mapping analysis. However, this method makes the simple assumption that the effect of a drug treatment is similar to knocking down its single target gene. Since many compounds can bind multiple targets, the pairwise mapping ignores the combined effects of multiple targets, and therefore fails to detect many potential targets of the compounds.

Results
We propose an algorithm to find sets of gene knock-downs that induce gene expression changes similar to a drug treatment. Assuming that the effects of gene knock-downs are additive, we first propose a novel bipartite block-wise sparse multi-task learning model~(BBS-MTL) for multi-target drug repositioning that overcomes the restrictive assumptions of connectivity mapping analysis. We then develop a BBSS-MTL model by integrating the drug structure information into BBS-MTL. By applying our approaches to five datasets generated from different cancer cell lines, we can predict potential drug targets more accurately than the simple pairwise connectivity mapping analysis.

Paper ID: 46

Accurate detection of time delays and directional interactions based on time series from complex dynamical systems

Huanfei Ma[1] and Wei Lin[2]
[1]Soochow University, China
[2]Fudan University, China

Data-based and model-free accurate identication of intrinsic time delays and directional interactions is an extremely challenging problem in complex dynamical systems and their networks reconstruction. A model-free method with new scores is proposed to be generally capable of

detecting

single, multiple, and distributed time delays. The method is applicable not only to mutually interacting dynamical variables but also to self-interacting variables in a time delayed feedback loop. Validation of the method is carried out using physical, biological and ecological models and real data sets. This method, as well as its application requirements, is expected to be applicable to ascertaining and quantifying subtle interactions (e.g., causation) in complex systems arising from a broad range of disciplines.

Paper ID: 47

Identification of novel drug targets for Diamond Black-fan Anemia (DBA) based on RPS19 gene mutation, using protein-protein interaction network

Abbas Khan[1], Arif Ali[1], Muhammad Junaid[1], William Cs Cho[2] and Dong-Qing Wei[1]
[1]Shanghai Jiao Tong University, China
[2]Queen Elizabeth Hospital, Hong Kong

Diamond-Blackfan anemia (DBA) is the most common haploinsufficiency in infant children. This study was conducted to identify novel therapeutic signatures during DBA and discover their mechanisms. The gene expression dataset of GSE14335 were downloaded from GEO database, containing 6 normal and 4 diseased samples. The gene ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes pathway (KEGG) enrichment analyses were performed, and protein–protein interaction (PPI) network of the differentially expressed genes (DEGs) was constructed by Cytoscape software. A total of 607 DEGs were identified in DBA, including 433 up-regulated genes and 174 downregulated genes. GO analysis results showed that up-regulated DEGs were significantly enriched in biological processes (BP), negative regulation of transcription from RNA polymerase II promoter, chemo taxis, inflammatory response, immune response, positive regulation of cell proliferation, negative regulation of cell proliferation, response to mechanical stimulus, positive regulation of cell migration, response to lipopolysaccharide, and defence response. KEGG pathway analysis revealed the TNF signalling pathway, Osteoclast differentiation, Chemokine signalling pathway, Cytokine -cytokine receptor interaction, Rheumatoid arthritis, Biosynthesis of amino acids, Biosynthesis of antibiotics and Glycine, serine and threonine metabolism. The top 10 hub genes, AKT1, IL6, NFKB1, STAT3, STAT1, RAC1, EGR1, IL8, RELA, RAC3, mTOR and CCR2 were identified from the PPI network, and sub-networks. In conclusion, the present study indicated that the identified DEGs and hub genes promote our understanding of the molecular mechanisms underlying the development of DBA, and might be used as molecular targets and diagnostic biomarkers for the treatment of DBA.

Paper ID: 48

VCNet: vector-based gene co-expression network construction and its application to RNA-seq data

Zengmiao Wang[1], Huaying Fang[1], Nelson Leung-Sang Tang[2] and Minghua Deng[1]
[1]Peking University, China
[2]The Chinese University of Hong Kong, Hong Kong

Building gene co-expression network (GCN) from gene expression data is an important field of bioinformatic research. Nowadays, RNA-seq data provides high dimensional information to quantify gene expressions in term of read counts for individual exons of genes. We develop an algorithm called VCNet to construct GCN from RNA-seq data to overcome this dimensional problem. VCNet performs a new statistical hypothesis test based on the correlation matrix of a gene–gene pair using the Frobenius norm. The asymptotic distribution of the new test is obtained under the null model. Simulation studies demonstrate that VCNet outperforms SpliceNet and RNASeqNet for detecting edges of GCN. We also apply VCNet to two expression datasets from TCGA database: the normal breast tissue and kidney tumour tissue, and the results show that the GCNs constructed by VCNet contain more biologically meaningful interactions than existing methods.

Paper ID: 49

A novel control strategy for finding driver nodes to target control complex networks

Wei-Feng Guo[1], Shao-Wu Zhang[1], Tao Zeng[2] and Luonan Chen[2]
[1]Northwestern Polytechnical University, China
[2]Shanghai Institutes for Biological Sciences, CAS, China

It is of great theoretical interest and practical significance to study how to control a system by applying perturbations to only a few driver nodes. Recently, a hot topic of modern network researches is how to determine the driver nodes that allow the control of an entire network. However, in practice, to control a complex network, especially biological networks, 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). Compared to the general concept of controllability, we introduce a concept of the constrained target controllability (CTC) of complex networks, which concerns 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 the constrained control nodes. To efficiently investigate CTC of complex networks, we further design a novel graph-theoretic algorithm, called CTCA, to estimate the ability of a given network to control the targets by choosing driver nodes from the set of the constrained control nodes. We extensively evaluate the constrained target controllability of numerous real complex networks. The results indicate that the biological networks with higher average degree are easier to be controlled than the biological networks with lower average degree while the electronic networks with lower average degree, are easier to be controlled than the web networks with higher average degree. We also show that our CTCA can more efficiently produce driver nodes for target controlling the networks than the existing state-of-the-art methods. Moreover, we have applied our CTCA to analyze two expert-curate bio-molecular networks and compared to other state-of-the-art methods. The results illustrate that our CTCA can efficiently identify the proved drug targets and new potentials, according to the constrained controllability of those biological networks.

Paper ID: 50

On the Statistical Signicance of Protein Complex

Youfu Su, Can Zhao, Zheng Chen, Bo Tian and Zengyou He
Dalian University of Technology, China

Motivation: Statistical validation of predicted complexes is a fundamental issue in proteomics and bioinformatics. The target is to measure the statistical signicance of each predicted complex in terms of p-values. Surprisingly, this issue has not received much attention in the literature. To our knowledge, only a few research eorts have been made towards this direction.

Result: In this article, we propose a novel method for calculating the p-value of a predicted complex. The null hypothesis is that there is no dierence between the number of edges in target protein complex and that in the random null model. In addition, we assume that a true protein complex must be a connected subgraph. Based on this null hypothesis, we present an algorithm to compute the p-value of a given predicted complex. We test our method on ve public available data sets. The experimental results show that our method is superior to the state-of-the-art statistical signicance testing algorithms for predicted complex evaluation.

Paper ID: 51

Sample-pattern Identification by Integrating Multiple Omics Data

Qianqian Shi[1], Chuanchao Zhang[2] and Luonan Chen[1]
[1]Shanghai Institutes for Biological Sciences, CAS, China
[2]Wuhan University, China

Motivation: Integrating different omics profiles is a challenging task, which provides a comprehensive way to understand complex diseases in a multi-view manner. One key for such an integration is to extract intrinsic patterns in concordance with data structures, so as to discover consistent information across various data types even with noise pollution. Thus, we proposed a novel framework called "pattern fusion analysis" (PFA), which performs automated information alignment and bias correction, to fuse local sample-patterns (e.g., from each data type) into a global sample-pattern corresponding to phenotypes (e.g., across most data types). In particular, PFA can identify significant sample-patterns from different omics profiles by optimally adjusting the effects of each data type to the pat-terns, thereby alleviating the problems to process different platforms and different reliability levels of heterogeneous data.

Results: To validate the effectiveness of our method, we first tested PFA on various synthetic datasets, and found that PFA can not only capture the intrinsic sample clustering structures from the multi-omics data in contrast to the state-of-the-art methods, such as iClusterPlus, SNF and moCluster, but also provide an automatic weight-scheme to measure the corresponding contributions by data types or even samples. In addition, the computational results show that PFA can reveal shared and complementary sample-patterns across data types with distinct signal-to-noise ratios in Cancer Cell Line Encyclopedia (CCLE) data sets, and outperforms over other works at identifying clinically distinct cancer subtypes in The Cancer Genome Atlas (TCGA) data sets.

Paper ID: 53

Functional Enrichment Analysis based on Long-noncoding RNA Associations

Tun-Wen Pai and Kuo-Sheng Hung
National Taiwan Ocean University, Taiwan

Hypoxia affects zebrafish embryo central neuron system (CNS) development through the expression of survivin (birc5) family. To discover regulation mechanisms in biological systems, RNA-seq of knock-down survivin zebrafish embryos was performed to obtain differentially expressed genes for functional enrichment analysis. In addition to use mRNAs with significant changes in traditional approaches, long noncoding RNAs associated genes were suggested to be involved for comprehensive investigation. Compared to traditional GO term and KEGG pathway enrichment analysis, using combination of differentially expressed mRNAs and long noncoding RNA (lncRNA) associated genes could identify several additional important GO terms and KEGG pathways. Furthermore, several lncRNA associated genes related to neuron development function could be retrieved in this study. It is also discovered from additionally identified pathways that FoxO and MAPK signaling pathways play an important role in connecting to apoptosis regulated by survivin. We demonstrated that incorporating differentially expressed lncRNAs associated genes outperformed traditional enrichment analysis for effective biological functional interpretations.

Paper ID: 54

Finding correlated pattern via high order matching for multiple sourced biological data

Xi Yang, Guoqiang Han and Hongmin Cai
South China University of Technology, China

The emerging multi-dimensional genomic data poses new challenges in data analysis. Finding correlated patterns from multiple sourced biological data is useful in understanding potential interacting relationships between the multi-modal genomic data. Multi-dimensional genomic data contain multiple genomic data types. Different types of genomic data have different scales and units. These data cannot be simply aggregated for analysis. To address this issue, a correlated pattern discovery model incorporating the prior knowledge is proposed. A tensor similarity is used to measure the correlation of common patterns. The model is combined with the prior knowledge. The expression form of prior knowledge is transformed as the constraints of prior knowledge. Efficient numerical solution is designed and analyzed. The proposed method is shown to perform robustly and effectively for both simulated data and real biological data. We conducted experiments on five real cancer datasets to reveal cancer subtypes. A survival analysis on the found subtypes confirmed the effectiveness of the model. This method is meaningful for doctors to realize personalized diagnosis and treatment of cancer and other diseases.

Paper ID: 56

The computational perspective for the HIV-1 entry inhibitor Maraviroc

Liang Yu and Luonan Chen
Shanghai Institutes for Biological Sciences, CAS, China

Maraviroc, as the unique HIV-1 entry inhibitor, has a side effect to harm the liver of patient discovered currently. Surely, patients will care all side effects of Maraviroc. The crystal structure of Maraviroc in complex with CCR5 opens the avenue to discover other side effects of Maraviroc and further discover other natural HIV-1 entry inhibitors from the panel of old drugs. In this paper, we first computationally prove that Maraviroc may complex with 20 enzymes as similar as that Maraviroc in complex with CCR5 and the value of minimal free energy (MFE) for Maraviroc binding to each of these 20 enzymes is less than the value of MFE for Maraviroc binding to CYP3A4. And we may further infer that Maraviroc have multiple side effects following from the roles of these 20 enzymes. As the result, we select azithromycin to replace Maraviroc as the HIV-1 entry inhibitor. We prove that it is good enough for binding to CCR5, although it is not stronger than Maraviroc. However, it is also weaker than Maraviroc for binding to the other GPCRs or human enzymes. In the sense of computer-aided drug design, azithromycin is a good HIV-1 entry inhibitor. It is worthy to be confirmed using wet lab.

Paper ID: 57

A novel method for identifying potential disease-related miRNAs via disease-miRNA-target heterogeneous network

Liang Ding, Minghui Wang, Dongdong Sun and Ao Li
University of Science and Technology of China, China

MicroRNAs (miRNAs), as a kind of important small endogenous single-stranded non-coding RNAs, play critical roles in a large number of human diseases. However, the currently known experimental verification of the disease-miRNA associations are still rare and experimental identification is time-consuming and labor-intensive. Accordingly, identifying potential disease-related miRNAs to help people understand the pathogenesis of complex diseases has become a hot topic. In this study, we take advantage of known disease-miRNA associations combined with a large number of experimentally validated miRNA-target associations, and further develop a novel disease-miRNA-target heterogeneous network for identifying disease-related miRNAs. The leave-one-out cross validation experiment and several statistical measures demonstrate that our method can effectively identify potential disease-related miRNAs. Furthermore, the good predictive performance of 15 common diseases and the manually confirmed analyses of top 30 candidates of hepatocellular carcinoma, ovarian neoplasms and breast neoplasms further provide convincing evidence of the practical ability of our method. The source code implemented by our method can be downloaded from the Github: https://github.com/USTC-HIlab/DMTHNDM.

Paper ID: 58

Coarse-grained core endogenous network predicts major pancreatic developmental lineages

Junqiang Wang[1], Ruoshi Yuan[1], Xiaomei Zhu[2] and Ping Ao[1,2]
[1]*Shanghai Jiao Tong University, China*
[2]*Shanghai University, China*

Pancreas development has attracted substantial research attention driven by two major diseases: diabetes and pancreatic cancer, which in turn has made it one of the best understood organogenesis model. However, till now, it is still unsatisfactory in understanding the mechanism that ensure the pancreatic cells differentiate in a precise order into different lineages during development. A core endogenous network for early pancreas development was constructed by summarizing the biochemical molecular knowledge of gene regulation. The multi-stability feature of the endogenous network was obtained after formulating the network into a set of coarse-grained ordinary differentiation equations. All the known major pancreatic cell types during early development are predicted by the stable states in the endogenous network. The adaptive landscape topology was also inferred from the endogenous network. Both the stepwise differentiation manner of most cells and the direct differentiation manner of the first wave endocrine cells are explained by the adaptive landscape. These two distinct differentiation strategies are reconciled in the adaptive landscape. In conclusion, the coarse-grained endogenous network reveals a relatively complete cell differentiation hierarchy in early pancreas development and provides insights into understanding the heterogeneity inherited in the developmental process.

Paper ID: 59

A branch point on differentiation trajectory is the bifurcating event revealed by dynamic network biomarker analysis of single-cell data

Ziwei Chen[1], Xiangqi Bai[1], Liang Ma[2], Xiawei Wang[3], Xiuqin Liu[4], Yuting Liu[5], Luonan Chen[6] and Lin Wan[1]

[1]*Academy of Mathematics and Systems Science, CAS, China*
[2]*Beijing Institute of Genomics, CAS, China*
[3]*University of Wisconsin-Madison, USA*
[4]*University of Science and Technology Beijing, China*
[5]*Beijing Jiaotong University, China*
[6]*Shanghai Institutes for Biological Sciences, CAS, China*

The advance in single-cell profiling technologies and the development in computational algorithms provide the opportunity to construct pseudo temporal trajectory with branching point of cellular development. On the other hand, theories such as dynamical network biomarkers (DNB) theory have been recently proposed to characterize the pre-transition state in real world dynamic systems. Few studies have validated whether the branch point identified in pseudo time is the critical point in real dynamic system. In this study, the dynamical behavior of the branching point on the pseudo trajectory has been investigated. We first reconstruct pseudo temporal trajectories with bifurcation in both single-cell mass cytometry data and single-cell RNA-seq(scRNA-seq) data by Wishbone algorithm. The cells which ordered according to the trajectory are then divided into even bins. We next statistically analyze the dynamical properties of markers/genes of cells within each bin. Finally the DNB theory is applied to assess the validity of branching event on the pseudo trajectories. Our results demonstrate that the branching point which recovered by Wishbone algorithm is confirmed as a transition state in cell differentiation process by the DNB theory. Furthermore, we find that an appropriate group of DNB will amplify the signal of critical event as defined in DNB theory. Our study provides biological insights on the pseudo trajectory with branch point in a dynamical view

and also indicates that DNB theory may serve as a benchmark to check the validity of the branch point.

Paper ID: 61

Block Spectral Clustering And Domains Association Using Signed Graphs

Liu Ye[1], Michael Ng[1] and Stephen Wu[2]
[1]Hong Kong Baptist University, Hong Kong
[2]The Institute of Statistical Mathematics, Japan

Multi-domain clustering have attracted research's attention in bioinformatics, because clustering gene, proteins and cells from different resource can give people a more global and accurate understanding of biological phenomenon. However, most of the existing multi-domain clustering algorithm seek a common module of all the domains, limiting their applicability to homogeneous environment. Thus, they assume the similarity weight between objects are positive, while it is desirable to allow the relationship labeled with negative weights. What's more, in reality, the consistency between different domains may differ, information from related domains can help to get more accurate partition. In this paper, we proposed a unsupervised block spectral clustering method with domain association for signed graph. Our idea is to construct a signed block Laplacian matrix with consistency of domain for multiple graph, and using its eigenvector to cluster and update the consistency among different domains simultaneously. Experiment results on synthetic data and real data demonstrate clustering accuracy and property of domain association.

Paper ID: 63

Fixation probability of a beneficial mutation conferring decreased generation time in changing environments

Fangshu Cui and Bo Yuan
Shanghai Jiao Tong University, China

The adaptation of populations relies on the fixation probability of new beneficial mutations. Further, the adaptation in variable environments has drawn great attentions. Here, we extend the classical result of the fixation probability of beneficial mutations obtained by Haldane, and estimate the fixation probability of a beneficial mutation with a reduced generation time in a changing environment. Assuming that the selective advantage is very small, we concentrate all the changing factors of environment on a single quantity: effective selective advantage. Using a time-dependent branching process, we get the analytic approximation for the fixation probability of beneficial mutations that decrease the generation time. Then, we apply this approximation to two interesting biological cases. In these two instances, this approximation we obtained is in good agreement with the exact value, which shows that our result is effective.

Paper ID: 64

Inference of crosstalk effects between DNA methylation and lncRNA regulation in NSCLC

Binhua Tang
Hohai University, China

Intercellular crosstalk effects between DNA methylation and lncRNA regulation remain elusive in lung carcinoma epigenetics. We present an application toolkit MetLnc in integration and annotation for group-wise NSCLC tissue-based DNA methylation and lncRNA profiling resources, thus to comprehensively analyze differentially methylated loci and lncRNAs through deep interrogation. Together with multiple analytic functions, MetLnc acts as a comprehensive approach on epigenetic omics integration and interrogation. Via the benchmark with group-wise NSCLC tissue profiling and TCGA cohort resources, we interrogate differentially methylated CpG loci and lncRNAs as meaningful clues for inferring crosstalk effects between DNA methylation and lncRNA regulation; together we conclude with investigated biomarkers for further epigenetics and clinical trial studies.

Paper ID: 65

Feedback regulation in a stem cell model with acute myeloid leukaemia

Jianfeng Jiao, Min Luo and Ruiqi Wang
Shanhai University, China

The haematopoietic lineages with leukaemia lineages which is tightly controlled by negative feedback inhibition of end-product is considered in this paper. 100 years ago, leukemia has been found. However, up to now, the exact cause is unknown, and many factors are thought to be associated with the pathogenesis of leukemia. Nevertheless it is very necessary to continue the profound study of the pathogenesis of leukemia. Here, we propose a new mathematical model which include a negative feedback control from the terminally differentiated cells of hematopoietic lineages to describe what the regulatory mechanisms of leukemia are by a set of ordinary differential equations. Afterwards, we carried out a detailed kinetic bifurcation analysis of the model, and obtained some useful results.

Paper ID: 68

Transcriptome and Metabolic Flux Analysis Reveal Shift of Metabolic Patterns during Rice Grain Development

Fangzhou Shen[1], Xuting Wu[2], Luoxi Shi[1], Yangmin Chen[1], Zhuo Wang[1], Xuan Li[2]
[1]Shanghai Jiao Tong University, China
[2]Institute of Plant Physiology and Ecology, CAS, China

Rice (Oryza sativa) is one of the most important grain crops, which serves as food source for nearly half of the world population. The study of rice development process as well as related strategies for production has made significant progress. However, the comprehensive study on development of different rice tissues at both transcriptome and metabolic flux level across different stages was lacked. In this study, by integrating transcriptome data of different rice tissues with a genome-scale rice metabolic model, we generated tissue-specific models and investigated the shift of metabolic patterns, and the discrepancy between transcriptome and metabolism level. We found although the

flux patterns are not very similar with the gene expression pattern, the tissues at booting stage and mature grain stage can be separately clustered by primary metabolism at either level. While the gene expression and flux distribution of secondary metabolism are more diverse across tissues and stages. The critical rate-limiting reactions and pathways were also identified. We compared the pairs of samples of the same tissues at different stages or the different tissues at same stage to uncover the difference between transcription level and metabolism level. Three compared pairs showed there are more altered pathways at gene expression level than metabolic level, which indicate the metabolism is more robust to reflect the phenotype. In conclusion, the tissue-specific models revealed more detail metabolic pattern shift among different tissues and stages, which is of great significance to uncover mechanism of rice grain development and further improve production and quality of rice.

Paper ID: 69

A Draft Genome of Myospalax baileyi and its Annotation

Liang Li[1], Zhanyu Wang[1], Wei Feng[1], Peng Shi[2] and Lei Li[1]
[1]Academy of Mathematics and Systems Science, CAS, China
[2]Kunming Institute of Zoology, CAS, China

The plateau zokor (Myospalax baileyi) is a burrowing rodent found in Qinghai-Tibet Plateau. It has undergone a long process of evolution to live in such an extreme environment. Using the genome assembler BAUM we developed recently, we obtained a high-quality draft genome sequence of the plateau zokor based on the unpublished sequencing data provided by Kunming Institute of Zoology. The statistics and evaluation of the genome assembly were reported. The BLAST results of the assembled scaffolds versus the mouse genome revealed the detailed genome rearrangements. Based on the assembly, we annotated the zokor genome in two ways. First, we performed ab initio gene prediction by GENSCAN, GlimmerHMM, and Augustus. Second, using an expression data set published earlier, we reconstructed the multi-tissue transcriptome by two reference-based methods: HISAT2 and StringTie. Besides, mapping the reads of Zokor transcripts to the mouse and rat genome shows that the genes of plateau zokor are quite different from those of the other two rodents. The draft genome and its annotation provides a basis for unravelling the genetic adaptions to the plateau zokor's extreme living environment.

Paper ID: 70

Large-scale determination and characterization of cell type-specific regulatory elements in the human genome

Can Wang and Shihua Zhang
Academy of Mathematics and Systems Science, CAS, China

Histone modifications have been widely elucidated to play vital roles in transcription regulation and cell identity. The Roadmap Epigenomics Consortium generated a reference catalogue of several key histone modifications across >100s of human cell types and tissues. Decoding these epigenomes into functional regulatory elements is a challenging task in computational biology. To this end, we

adopted a differential chromatin modification analysis framework to comprehensively determine and characterize cell type-specific regulatory elements (CSREs) and their histone modification codes in the human epigenomes of five histone modifications across 127 tissues or cell types. The CSREs show significant relevance with cell type-specific biological functions and diseases and cell identity. Clustering of CSREs with their specificity signals reveals diverse histone codes, demonstrating the diversity of functional roles of CSREs within the same cell or tissue. Last but not least, dynamics of CSREs from close cell types can give a detailed view of developmental process such as normal tissue development and cancer occurrence.

Paper ID: 71

Hierarchical combinatorial deep learning architecture for pancreas segmentation for medical computed tomography cancer images

Min Fu[1], Qiuhua Liu[1], Yaobin Ou[1], Yupei Zhao[2,3] and Xinqi Gong[1]
[1]Renmin University of China, China
[2]Chinese Academy of Medical Sciences, China
[3]Peking Union Medical College, China

Computational efficiently recognition and segmentation of concerning objects from medical images are very important for diagnosis and treatments. This paper adopts a new architecture workflow based on an advanced deep network with rich convolutional features to extract the pancreas from computed tomography (CT) images. Specifically, pancreas segmentation is formalized as an edge detective and pattern recognition problem, where the ground-truth tells the class of a given pixel, positive (pancreas) or negative (non-pancreas). Each pixel is described by some features containing information of the CT scanning pictures. Based on the generated features, our architecture workflow is employed to perform pixel classification and pancreas segmentation. Since the great diversity in shape, size, and location among the pancreas of different people, especially the different kinds of pancreas cancers and different phases of cancers, multistage methods are adopted to acquire more information and improve the accuracy of segmentation. Our model is validated on a set of CT data. The result shows that the rich architecture workflow performs well in pancreas segmentation.

Paper ID: 72

PPI network analyses of human WD40 protein family systematically reveal their tendency to assemble complexes and facilitate the complex predictions

Xu-Dong Zou, Ke An, Yun-Dong Wu and Zhi-Qiang Ye
Peking University Shenzhen Graduate School, China

Background

WD40 proteins constitute one of the largest families in eukaryotes, and more than 260 human genes belong to this family. Although previous studies have speculated that WD40 proteins should participate frequently in the protein-protein interactions (PPI) based on the available crystal structures, this inference requires confirmation quantitatively at the systems level. Many WD40 protein-associated complexes have been reported to be responsible for diverse and key cellular

processes, but the roles that WD40 proteins play in organizing cellular networks remains unclear. In addition, studies on their network characteristics may provide clues to predict putative complexes and to prioritize certain specific WD40 proteins for further detailed investigations, and are thus highly required. To date, no systematic analysis has addressed these issues from a network perspective.

Results

We built two human PPI networks using PPI data sets with different confidence levels, and applied network-based methods to study the topological properties of human WD40 proteins systematically. Comparing with non-WD40 proteins, we have found three distinct characteristics: (i) they tend to be hubs; (ii) they are prone to locate near the global center; (iii) WD40 hubs are inclined to be intramodular. Using these network features, certain WD40 proteins are specifically reviewed and can be prioritized for further studies. Moreover, we have predicted 1674 potential complexes for 150 human WD40 proteins. The indirect evaluation based on co-expression scores has shown that the predicted complexes are evidently different from the random data set, but are very similar to the validated reference complexes, indicating our predictions based on the network are reliable.

Conclusion

At the systems level but not sporadic examples' level, our results have confirmed that WD40 proteins should tend to be intramodular hubs, responsible for assembling complexes in organizing the human PPI network. Our analyses also provided network-based clues for prioritizing WD40 proteins and for predicting complexes.

Paper ID: 73

Predicting disease by edge network analysis

Xiangtian Yu, Tao Zeng and Luonan Chen
Institute of Biochemistry and Cell Biology, CAS, China

Predicting pre-disease state or tipping point just before irreversible deterioration is a difficult task. Edge network analysis opens a new way to study this problem by exploring rich dynamical and high-dimensional information of omics data. Although theoretically edge network analysis has the ability to identify the pre-disease state or the tipping point during the disease progression, it requires multiple samples for such prediction on each individual, which are generally not available in clinical practice, thus limiting its wide applications in personalized medicine. In this work to overcome this problem, we propose an improved edge network analysis to identify the pre-disease state of each individual based on a single sample. In particular, this method can identify the common edge-biomarkers from individual-specific biomarkers, which are used for the disease prediction. To demonstrate the effectiveness, our method was applied to the analysis on omics data of H3N2 and H1N1 cohorts, and successfully detected early-warning signals or tipping points of the influenza infection for each individual in an accurate manner. In addition, the proposed method not only found the new individual-specific biomarkers but also recovered the common biomarkers of influenza infection reported from previous works.

Paper ID: 74

Prediction of drug response with a dual-layer network model

Suyun Huang and Xingming Zhao
Tongji University, China

Identifying the response of a cancer patient to a particular therapeutic agent is critical in drug discovery and will significantly facilitates the development of personalized medicine. The publicly available drug response profiles across cell lines provide an alternative way for predicting the response of cancer drugs. In this work, we propose a dual-layer network model to predict drug response based on large-scale cell line experiments. With the Cancer Cell Line Encyclopedia (CCLE), Genomics of Drug Sensitivity in Cancer (GDSC) and Cancer Therapeutic Response Portal (CTRP) datasets as benchmark datasets, our proposed dual-layer network model outperforms other existing popular approaches and identify some novel indications of known drugs for cancer.

Paper ID: 76

m6A-Driver: Identifying Context-Specific mRNA m6A Methylation-Driven Gene Interaction Networks

Song-Yao Zhang[1], Shao-Wu Zhang[1], Lian Liu[1], Jia Meng[2] and Yufei Huang[3]
[1]*Northwestern Polytechnical University, China*
[2]*Xi'an Jiaotong-Liverpool University, China*
[3]*University of Texas at San Antonio, USA*

As the most prevalent mammalian mRNA epigenetic modification, N6-methyladenosine(m6A) has been shown to possess important post-transcriptional regulatory functions. However, the regulatory mechanisms and functional circuits of m6A are still largely elusive. To help unveil the regulatory circuitry mediated by mRNA m6A methylation, we develop here m6A-Driver, an algorithm for predicting m6A-driven genes and associated networks, whose functional interactions are likely to be actively modulated by m6A methylation under a specific condition. Specifically, m6A-Driver integrates the PPI network and the predicted differential m6A methylation sites from methylated RNA immunoprecipitation sequencing (MeRIP-Seq) data using a Random Walk with Restart (RWR) algorithm and then builds a consensus m6A-driven network of m6A-driven genes. To evaluate the performance, we applied m6A-Driver to build the context-specific m6A-driven networks for 4 known m6A (de)methylases, i.e., FTO, METTL3, METTL14 and WTAP. Our results suggest that m6A-Driver can robustly and efficiently identify m6A-driven genes that are functionally more enriched and associated with higher degree of differential expression than differential m6A methylated genes. Pathway analysis of the constructed context-specific m6A-driven gene networks further revealed the regulatory circuitry underlying the dynamic interplays between the methyltransferases and demethylase at the epitranscriptomic layer of gene regulation

Paper ID: 77

A novel peptide specifically binding to VEGF receptor suppresses angiogenesis in vitro and in vivo

Bifang He, Lin Ning and Jian Huang
University of Electronic Science and Technology of China, China

Vascular endothelial growth factor (VEGF), one of the most important angiogenic factors, plays an essential role in both physiological and pathological angiogenesis through binding to VEGF receptors (VEGFRs). Here we report a novel peptide designated HRHTKQRHTALH (peptide HRH), which was isolated from the Ph.D. -12 phage display library using VEGFR-Fc fusion protein as the bait. This peptide was found to dose-dependently inhibit the proliferation of human umbilical vein endothelial cells stimulated by VEGF. The anti-angiogenesis effect of the HRH peptide was further confirmed in vivo using the chick chorioallantoic membrane assay, which was also dose-dependent. Besides, peptide HRH was proved to inhibit corneal neovascularization in an alkali-burnt rat corneal model and a suture-induced rat corneal model. Taken together, these findings suggest that the HRH peptide can inhibit angiogenesis both in vitro and in vivo. Consequently, the HRHTKQRHTALH peptide might be a promising lead peptide for the development of potential angiogenic inhibitors.

Paper ID: 79

A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation

Chunhe Li[1], Tian Hong[2], Chiu-Ho Webb[2], Heather Karner[2], Sha Sun[2] and Qing Nie[2]
[1]Fudan University, China
[2]University of California, Irvine, USA

X-chromosome inactivation (XCI) is the mammalian dosage compensation strategy for balancing sex chromosome content between females and males. While works exist on initiation of symmetric breaking, the underlying allelic choice mechanisms and dynamic regulation responsible for the asymmetric fate determination of XCI remain elusive. Here we combine mathematical modeling and experimental data to examine the mechanism of XCI fate decision by analyzing the signaling regulatory circuit associated with long noncoding RNAs (lncRNAs) involved in XCI. We describe three plausible gene network models that incorporate features of lncRNAs in their localized actions and rapid transcriptional turnovers. In particular, we show experimentally that Jpx (a lncRNA) is transcribed biallelically, escapes XCI, and is asymmetrically dispersed between two X's. Subjecting Jpx to our test of model predictions against previous experimental observations, we identify that a self-enhanced transport feedback mechanism is critical to XCI fate decision. In addition, the analysis indicates that an ultrasensitive response of Jpx signal on CTCF is important in this mechanism. Overall, our combined modeling and experimental data suggest that the self-enhanced transport regulation based on allele- specific nature of lncRNAs and their temporal dynamics provides a robust and novel mechanism for bi- directional fate decisions in critical developmental processes.

Paper ID : 80

MetaMHCpan, A Meta Approach for Pan-Specific MHC Peptide Binding Prediction

Shanfeng Zhu
Fudan University, China

Recent computational approaches in bioinformatics can achieve high performance, by which they can be a powerful support for performing real biological experiments, making biologists pay more attention to bioinformatics than before. In immunology, predicting peptides which can bind to MHC alleles is an important task, being tackled by many computational approaches. However, this situation causes a serious problem for immunologists to select the appropriate method to be used in bioinformatics. To overcome this problem, we develop an ensemble prediction-based Web server, which we call MetaMHCpan, consisting of two parts: MetaMHCIpan and MetaMHCIIpan, for predicting peptides which can bind MHC-I and MHC-II, respectively. MetaMHCIpan and MetaMHCIIpan use two (MHC2SKpan and LApan) and four (TEPITOPEpan, MHC2SKpan, LApan, and MHC2MIL) existing predictors, respectively. MetaMHCpan is available at http://datamining-iip.fudan.edu.cn/MetaMHCpan/index.php/pages/view/info.

Paper ID: 82

Quantifying direct dependencies in biological networks by multiscale association analysis

Jifan Shi[1], Luonan Chen[2], Tiejun Li[1], Juan Zhao[2] and Xiaoping Liu[2]
[1]Peking University, China
[2]Shanghai Institutes for Biological Sciences, CAS, China

Partial correlation (PC) or conditional mutual information (CMI) is widely used in detecting direct dependencies between the observed variables in biological networks by eliminating indirect correlations/associations, but it fails whenever there are some strong correlations in a network. In this paper, we theoretically develop a multiscale association analysis to overcome this flaw. We propose a new measure, partial association (PA), based on the multiscale conditional mutual information. We show that linear PA and nonlinear PA have clear advantages over PC and CMI from both theoretical and computational aspects. Both simulated models and real omics datasets demonstrate that PA is superior to PC and CMI in terms of accuracy, and is a powerful tool to identify the direct associations or reconstruct molecular networks based on the observed data. Survival and functional analyses of the hub genes in the gene networks reconstructed from TCGA data for different cancers also validated the effectiveness of our method.

Paper ID: 84

Proteomics toolbox for profiling cancer signaling - Toward precision medicine

Ruijun Tian
Southern University of Science and Technology, China

Cell-cell interactions are often mediated by protein molecules expressed by one cell that are recognized by specific receptors present on neighboring cells. This intercellular communication

activates specific signaling pathways through the induction of dynamic posttranslational modifications such as phosphorylation and protein-protein interactions. Mass spectrometry-based proteomics have been proven to be a robust approach for characterizing dynamic protein phosphorylation and interaction on a global scale. However, proteomic studies of signaling networks under conditions with physiological cell-cell contact and with limited sample access have been challenging. In this talk, I will present our recent development of unbiased proteomic approaches toward this end. Our work established several generally applicable and combinatorial strategies for studying intercellular cancer signaling and provided valuable resources for systematically understanding dysfunctional signaling networks in tumor microenvironment in a context closer to physiological condition than was previously possible.

Paper ID: 85

Single-cell transcriptome analysis reveals widespread monoallelic gene expression in individual rice mesophyll cells

Wenfeng Qian
Institute of Genetics and Developmental Biology, CAS, China

Monoallelic gene expression refers to the phenomenon that all transcripts of a gene in a cell are expressed from only one of the two alleles in a diploid organism. Although monoallelic gene expression has been occasionally reported with bulk transcriptome analysis in plants, how prevalent it is in individual plant cells remains unknown. Here, we developed a single-cell RNA-seq protocol in rice and investigated allelic expression patterns in mesophyll cells of indica (93-11) and japonica (Nipponbare) inbred lines, as well as their F1 reciprocal hybrids. We observed pervasive monoallelic gene expression in individual mesophyll cells, which could be largely explained by stochastic and independent transcription of two alleles. By contrast, two mechanisms that were proposed previously based on bulk transcriptome analyses, parent-of-origin effects and allelic repression, were not well supported by our data. Furthermore, monoallelically expressed genes exhibited a number of characteristics, such as lower expression levels, narrower H3K4me3/H3K9ac/H3K27me3 peaks, and larger expression divergences between 93-11 and Nipponbare. Taken together, the development of a single-cell RNA-seq protocol in this study offers us an excellent opportunity to investigate the origins and prevalence of monoallelic gene expression in plant cells.

Paper ID: 86

Interrogating RNA interactomes

Qiangfeng Zhang
Tsinghua University, China

RNA molecules in the cell never exist alone. From their birth to death, they interact with many different molecules, including proteins, DNA and other RNAs. These interactions are essential to understanding the biological functions and molecular mechanisms of both messenger RNAs (mRNAs) and noncoding RNAs (ncRNAs). Here we report our recent progress in revealing RNA-

protein and RNA-RNA interactomes in different cell lines. We compared different RNA interactome data sets and and found limited overlaps in interactions resolved by different techniques and in different cell lines. It may suggest technology preference and dynamic natures of RRIs. We also analyzed the RNA interaction network characteristics and found it tends to be scale-free, small-world, hierarchical, and modular.

Paper ID: 87

Tumor evolution and precision cancer medicine

Jiguang Wang
Hong Kong University of Science and Technology, Hong Kong

Precision medicine in cancer proposes that genomic characterization of tumors can inform personalized targeted therapies. However, this proposition is complicated by spatial and temporal heterogeneity. Here we study genomic and expression profiles from multisector or longitudinal specimens from patients with glioblastoma (GBM). Using bulk and single-cell data, we find that samples from the same tumor mass share genomic and expression signatures, whereas geographically separated, multifocal tumors and/or long-term recurrent tumors are seeded from different clones. Chemical screening of patient-derived glioma cells (PDCs) shows that therapeutic response is associated with genetic similarity, and multifocal tumors that are enriched with PIK3CA mutations have a heterogeneous drug-response pattern. We show that targeting truncal events is more efficacious than targeting private events in reducing the tumor burden. In summary, this work demonstrates that evolutionary inference from integrated genomic analysis in multisector biopsies can inform targeted therapeutic interventions for patients with GBM.

Paper ID: 89

Identification of PTMs in regulating autophagy

Wankun Deng and Yu Xue
Huazhong University of Science and Technology, China

Autophagy is a highly conserved process for degrading cytoplasmic contents, determines cell survival or death, and regulates the cellular homeostasis. Besides core ATG proteins, numerous regulators together with various post-translational modifications (PTMs) are also involved in autophagy. Recent studies demonstrated that the dysregulation of macroautophagy/autophagy is involved in human diseases such as cancers and neurodegenerative disorders. Thus, autophagy has become a promising therapeutic target for biomedical design. Here, we developed a database of The Autophagy, Necrosis, ApopTosis OrchestratorS (THANATOS, http://thanatos.biocuckoo.org), containing 191,543 proteins potentially associated with autophagy and cell death pathways in 164 eukaryotes. We performed an evolutionary analysis of core ATG genes, and observed that ATGs required for the autophagosome formation are highly conserved across eukaryotes. Further analyses revealed that known cancer genes and drug targets were over-represented in human autophagy proteins, which were significantly associated in a number of signaling pathways and human diseases. By re-constructing a human kinase-substrate phosphorylation network for core ATG proteins, our

results confirmed that phosphorylation play a critical role in regulating autophagy. Using this data resource, we performed a quantitative phosphoproteomic profiling to delineate the phosphorylation signalling networks regulated by 2 natural neuroprotective autophagy enhancers, corynoxine (Cory) and corynoxine B (Cory B). We developed a novel algorithm of in silico Kinome Activity Profiling (iKAP) to predict that Cory or Cory B potentially regulates different kinases. We discovered 2 kinases, MAP2K2/MEK2 (mitogen-activated protein kinase kinase 2) and PLK1 (polo-like kinase 1), to be potentially upregulated by Cory, whereas the siRNA-mediated knockdown of Map2k2 and Plk1 significantly inhibited Cory-induced autophagy. Furthermore, Cory promoted the clearance of Alzheimer disease-associated APP (amyloid beta [A4] precursor protein) and Parkinson disease-associated synuclein alpha (SNCA/α-synuclein) by enhancing autophagy, and these effects were dramatically diminished by the inhibition of the kinase activities of MAP2K2 and PLK1. Taken together, our study not only provided bioinformatics resources and approaches for analyzing PTMs in autophagy, but also identified the important role of MAP2K2 and PLK1 in neuronal autophagy.

Paper ID: 90

Rethink the coding capacity of non-coding RNA: extensive translation of circRNA driven by RNA Methylation

Zefeng Wang
Partner Institute for Computational Biology, China

Thousands of circular RNAs (circRNAs) have been detected recently, but in vivo functions of most circRNAs are unclear. Previous studies showed that artificial circRNAs consisting of internal ribosome entry site (IRES) can be translated in vitro, however no evidence to support in vivo translation of circRNA. Meanwhile human RNAs are extensively modified, with N6-methyladenosine (m6A) being the most abundant base modification. Here we find that m6A sites efficiently initiate protein translation from a circRNA in human cells. Consistently, consensus m6A motifs are enriched in circRNAs. We further demonstrate that such translation is enhanced by m6A writer (METTL3/14), decreased by m6A eraser (FTO), and upregulated under heat shock condition. Moreover, we find that m6A-mediated translation of circRNA is promoted effectively by EIF4G2. We computationally predict endogenous circRNAs with translation potential, and have verified that most predicted circular mRNAs are associated with polysome. We then confirmed the protein translation from an endogenous circular mRNA, which is reduced by FTO and decreased by knocking down of EIF4G2. Finally, we perform high-throughput sequencing to globally identify translatable circRNAs and discover many polysome associated circRNAs with m6A binding site. In addition, we found that adenosine methylation is not always required for circRNA translation, further expanded the coding potential of circRNA. Our study expands the coding landscape of human transcriptome, and suggests a role of circRNA-derived proteins in cellular responses to environmental stress.

Paper ID: 91

Edge-based Integration of Multi-omics Data to Predict Disease State

Rong Zeng
Shanghai Institute of Biochemistry and Cell Biology, CAS, China

We develop a novel method using edge information as features or edge-biomarkers to predict disease state. Specifically, this method transforms (time-varying or condition-dependent) molecular expression data into (relatively time-invariant or condition-independent) correlation data of each molecular pair on a single sample basis, thus achieving feature selection, classifier training and phenotype prediction on an edge or network level. The molecular pairs selected by a specific machine learning method are a new type of biomarkers called edge-biomarkers, which can accurately and stably diagnose disease state of each individual even with single sample data in contrast to the traditional molecule- or node-biomarkers. Due to the correlation nature of edge transformation, this method is feasible to construct molecular pairs from multi-omics data, resulting in edge-based integration (EI). We used this platform to evaluate diabetes risk and prognosis.

Paper ID: 92

Epitranscriptomic RNA Modifications: Regulations and Mechanisms

Yungui Yang
Beijing Institute of Genomics, CAS, China

Over 100 types of chemical modifications have been identified in various types of RNAs including non-coding RNA and mRNA, among which methylation is the most common modification. The N6-methyl-adenosine (m6A) and 5-methylcytosine (m5C) are the common and abundant internal modifications on mRNA molecules. The recent identification of m6A and m5C methyltransferase and m6A demethylases ALKBH5 and FTO, supports the reversibility of RNA methylation. Several m6A YTH-domain-containing proteins YTHDF1-3 and YTHDC1 and m5C binding protein ALYREF have been identified and regulate various mRNA processing, suggesting vital roles of RNA methylations in gene expression control. Depletion or inhibition of RNA methylations leads to dysregulate biological process, including lipid metabolism, spermatogenesis, neurogenesis, and haematopoietic stem cell specification. We will summarize recent progress in RNA methylations and discuss their potential biological significance in this conference.

Paper ID: 93

Critical behaviors of structural fluctuations in the native states of proteins

Wei Wang
Nanjing University, China

To achieve its biological functions, the structure of the native state of a protein must be susceptible enough to sense the signal and switch to another structure, but also be stable enough to warrant functional specificity and structural robustness. This means a coexistence of high susceptibility and stability for the protein around its native state, which is apparently competing since high

susceptibility implies large fluctuations and thus small stability in general, and vice versa. Does the balance of such competition result in a certain kind of critical behavior in proteins? Based on protein structural ensembles determined by NMR, we study the position fluctuations of residues by calculating distance-dependent correlations and conducting finite-size scaling analysis. The fluctuations exhibit high susceptibility and long-range correlations up to the protein sizes. The scaling relations between the correlations or susceptibility and protein sizes resemble those in other physical and biological systems near their critical points. These results indicate that, at the native states, motions of each residue are felt by every other one in the protein. We also find that proteins with larger susceptibility are more frequently observed in nature. Overall, our results suggest that the protein's native state is critical.