The 12th International Conference on Computational Systems Biology (ISB 2018)

August 18-21, 2018
Guiyang, China

Sponsors

贵州中科生态云大数据科技有限公司

ISB 2018 Schedule

Update: 2018/8/10 1:31 PM

August 18 Saturday	15:00-22:00	Registration (Hotel Lobby)	
	18:00-20:00	Reception	
	20:00-21:30	Board member meeting of ORSC-CSB (Room T2)	
August 19 Sunday	08:30-08:40	Opening Session (Room KH)	
	08:40-10:00	Plenary Session P1 (Room KH)	
	10:00-10:30	Coffee break	
	10:30-12:10	Highlight Session A1 (Room KH) Topic: Bioinformatics Paper IDs: 10, 47, 66, 71, 101 Chair: Shihua Zhang	Session B1 (Room T2) Topic: Systems Biology Paper IDs: 23, 28, 41, 57, 58 Chair: Jian Huang
	12:30-13:30	Lunch	
	14:00-15:40	Highlight Session A2 (Room KH) Topic: Systems Biology Paper IDs: 8, 9, 56, 69, 102 Chair: Yu Xue	Session B2 (Room T2) Topic: Bioinformatics Paper IDs: 6, 12, 22, 34, 35 Chair: Fengfeng Zhou
	15:40-16:20	Coffee break & Poster Session	
	16:20-18:00	Highlight Session A3 (Room KH) Topic: Proteomics Paper IDs: 3, 43, 55, 70, 103 Chair: Qiangfeng Zhang	Session B3 (Room T2) Topic: Systems Biology Paper IDs: 11, 24, 30, 32, 38 Chair: Zengyou He
	18:30-20:00	Banquet	
	20:30-21:30	Board member meeting of CSBMB-MSB (Room T2)	

ISB 2018 Schedule

Update: 2018/8/10 1:31 PM

August 20 Monday	08:30-10:30	Plenary Session P2 (Room KH)	
	10:30-10:50	Coffee break	
	10:50-12:30	Highlight Session A4 (Room KH) Topic: Systems Biology Paper IDs: 18, 40, 46, 72, 83 Chair: Anyuan Guo	Session B4 (Room T2) Topic: Diseases and Drugs Paper IDs: 16, 20, 31, 36, 61 Chair: Lei Li
	12:30-13:30	Lunch	
	14:00-15:40	Highlight Session A5 (Room KH) Topic: Network Biology Paper IDs: 81, 42, 49, 65, 67 Chair: Xianwen Ren	Session B5 (Room T2) Topic: Diseases and Drugs Paper IDs: 7, 14, 33, 39, 60 Chair: Jinzhi Lei
	15:40-16:20	Coffee break & Poster Session	
	16:20-18:00	Highlight Session A6 (Room KH) Topic: Systems Biology Paper IDs: 4, 44, 45, 75, 105 Chair: Yong Wang	Session B6 (Room T2) Topic: Proteomics Paper IDs: 13, 26, 37, 59, 62 Chair: Binqiang Liu
	18:30-20:00	Dinner	
August 21 Tuesday	08:00-13:00	Half day excursion in Qingyan Ancient Town. Departure at 8:00 from lobby.	

***** Subjects to revision based on further information

Room KH: Kendo Hall at 2nd floor (二楼剑道馆)

Room T2: Tianzi 2 Conference Room at 1st floor (一楼天字二号会议室)

ISB 2018 Program

August 18-21, Guiyang, China

August 18 (Saturday) Registration

15:00-22:00 Registration, Participants arrival in Guiyang, 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 (Tianzi 2 Conference Room at 1st floor 一楼天字二号会议室)

August 19 (Sunday) Technical sessions

08:30-08:40 Opening Session (Kendo Hall at 2nd floor 二楼剑道馆)
Chair: Luonan Chen

08:40-10:00 Plenary Session P1 (Kendo Hall at 2nd floor 二楼剑道馆)
Chair: Luonan Chen

08:40-09:20 *Deciphering developmental patterning with deep neural network*
Chao Tang[1,2,3]
[1]Center for Quantitative Biology, Peking University, Beijing 100871, China
[2]School of Physics, Peking University, Beijing 100871, China
[3]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

09:20-10:00 *Machine learning and complex networks for complex systems big data analysis and precision medicine*
Carlo Vittorio Cannistraci
Biomedical Cybernetics Group, Technical University Dresden, Germany

10:00-10:30 Coffee break

10:30-12:10 Highlight Session A1 (Kendo Hall at 2nd floor 二楼剑道馆)
Topic: Bioinformatics
Chair: Shihua Zhang

10:30-10:50 *GSCALite: A Web Server for Gene Set Cancer Analysis*
Anyuan Guo
Huazhong University of Science and Technology, China
Paper ID: 10

10:50-11:10 *Concept and application of differential network model in single-sample analysis of biological big data*
Tao Zheng
Key Laboratory of Systems Biology, CAS, China
Paper ID: 47

11:10-11:30 *TriVote, a highly accurate OMIC biomarker detection algorithm*
Fengfeng Zhou
Jilin University, China
Paper ID: 66

11:30-11:50 *The ultrafast and accurate mapping algorithm FANSe3: mapping a human whole genome sequencing dataset within one hour*
Gong Zhang[1,2,*], Yongjian Zhang[2] and Jingjie Jin[1]

[1]Jinan University, China
[2]Chi-Biotech Co. Ltd., China
Paper ID: 71

11:50-12:10 *EEG based Intelligent robot control*
Li Kun
School of Life Sciences, Guizhou Normal University
Paper ID: 101

10:30-12:10 Session B1 (Tianzi 2 Conference Room at 1st floor 一楼天字二号会议室)

Topic: Systems Biology
Chair: Jian Huang

10:30-10:50 *Extracting Predictors for Lung Adenocarcinoma Based on Granger Causality Test and Stepwise Character Selection*
Xuemeng Fan, Xuqing Tang*
Jiangnan University, China
Paper ID: 23

10:50-11:10 *Predicting stage-specific cancer related genes and their dynamic modules by integrating multiple datasets.*
Aouiche Chaima, Bolin Chen*, Xuequn Shang
Northwestern Polytechnical University, China
Paper ID: 28

11:10-11:30 *Generalized Gene Co-Expression Analysis via Subspace Clustering Using Low-Rank Representation*
Tongxin Wang[1], Jie Zhang[2] and Kun Huang[2,*]
[1]Indiana University Bloomington, USA
[2]Indiana University School of Medicine, USA
Paper ID: 41

11:30-11:50 *scLRTD : A novel low rank tensor decomposition method for single-cell RNA-sequencing data imputing*
Zhijie Ni, Dingjie Wang, Xiao Zheng and Xiufen Zou*
Wuhan University, China
Paper ID: 57

11:50-12:10 *Investigation of lipid metabolism dysregulation and the effects on immune microenvironment in pan-cancer using multiple omics data*
Yang Hao[1,2], Daixi Li[1,*], Yong Xu[1,2], Jian Ouyang[2], Yongkun Wang[1,2], Yuqi Zhang[1,2], Baoguo Li[1], Lu Xie[2,*] and Guangrong Qin[2,*]
[1]University of Shanghai for Science and Technology, China
[2]Shanghai Academy of Science and Technology, China
Paper ID: 58

12:30-13:30 Lunch

14:00-15:40 Highlight Session A2 (Kendo Hall at 2nd floor 二楼剑道馆)

Topic: Systems Biology

Chair: Yu Xue

14:00-14:20 *Data Science to Invent Digital Ag Solution*

Le Lv[1,*], and Jingdong Liu[2]

[1]Monsanto Biotech(Beijing) Research Center, China

[2]Monsanto Company, United States

Paper ID: 8

14:20-14:40 *Computational Design of Antiangiogenic Peptibody by Fusing Human IgG1 Fc Fragment and HRH Peptide: Structural Modeling, Energetic Analysis, and Dynamics Simulation of Its Binding Potency to VEGF Receptor*

Lin Ning, Zhongyan Li, Zhengya Bai, Shasha Hou, Bifang He, Jian Huang*, Peng Zhou*

Center for Information Biology, University of Electronic Science and technology of China, China

Paper ID: 9

14:40-15:00 *Profiling and functional analysis of circular RNAs in acute promyelocytic leukemia and their dynamic regulation during all-trans retinoic acid treatment*

Shufen Li, Yunlin Ma, Yun Tan, Xuefei Ma, Ming Zhao, Bing Chen, Rongsheng Zhang, Zhu Chen and Kankan Wang*

State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, China

Paper ID: 56

15:00-15:20 *De novo assembled individual genome doesn't show advantage against standard reference genome: a demonstration of Chinese Han Population*

Zhibiao Mai, Wanting Liu, Wen Ding and Gong Zhang*

Jinan University, China

Paper ID: 69

15:20-15:40 *基因测序数据处理的高性能计算挑战*

朱红

浪潮集团 高性能应用支持专家

Paper ID: 102

14:00-15:40 Session B2 (Tianzi 2 Conference Room at 1st floor 一楼天字二号会议室)

Topic: Bioinformatics

Chair: Fengfeng Zhou

14:00-14:20 *CrystalM: a multi-view fusion method for protein crystallization prediction*

Yubo Wang[1], Jijun Tang[2,*], Yu Dai[1], Fei Guo[1,*]

[1]Tianjin University, China

[2]University of South Carolina, US

Paper ID: 6

14:20-14:40 *Cross-species Data Classification by Domain Adaptation via Discriminative Heterogeneous Maximum Mean Discrepancy*

Limin Li* and Menglan Cai

Xi'an Jiaotong University, China

Paper ID: 12

14:40-15:00 *A deep learning framework for identifying essential proteins by integrating multiple sources of biological information*

Min Zeng[1], Min Li[1,*], Zhihui Fei[1], Fangxiang Wu[2], Yaohang Li[3], Yi Pan[4], Jianxin Wang[1]

[1]School of Information Science and Engineering, Central South University, China

[2]University of Saskatchewan, Canada

[3]Old Dominion University, United States

[4]Georgia State University, United States

Paper ID: 22

15:00-15:20 *Prioritizing type 2 diabetes genes by weighted PageRank on bilayer heterogeneous networks*

Haixia Shang, Zhiping Liu*

Shandong University, China

Paper ID: 34

15:20-15:40 *CNAPE: A Software for Copy Number Alteration Prediction from Gene Expression in Human Cancers*

Quanhua Mu, Jiguang Wang*

The Hong Kong University of Science and Technology, Hong Kong

Paper ID: 35

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

16:20-18:00 Highlight Session A3 (Kendo Hall at 2nd floor 二楼剑道馆)

Topic: Proteomics

Chair: Qiangfeng Zhang

16:20-16:40 *CHDOCK: A hierarchical docking approach for predicting the structure of homo-oligomeric complexes with Cn symmetry*

Yu-Meng Yan and Sheng-you Huang*

Huazhong University of Science and Technology, China

Paper ID: 3

16:40-17:00 *Discovery of the substrate of deubquitinase USP9X by quantitative proteomics*

Hu Zhou*, Xiangling Chen, Chengli Yu and Jing Gao*

Shanghai Institute of Materia Medica, CAS, China

Paper ID: 43

17:00-17:20 *Photocatalytic protein damage of silver nanoparticles circumvents bacterial stress-response and thus abolish multidrug resistance*

Tianyuan Shi, Qiuxia Wei, Gong Zhang*, Xuesong Sun* and Qing-Yu He*
Jinan University, China
Paper ID: 55

17:20-17:40 *Multifaceted stoichiometry control of bacterial operons revealed by data-independent acquisition mass spectrometry*
Jing Zhao[1], Hong Zhang[1], Bo Qin[2], Qingyu He[1], Rainer Nikolay[2], Christian Spahn[2] and Gong Zhang[1,*]
[1]Institute of Life and Health Engineering, Jinan University, China
[2]Institut für Medizinische Physik und Biophysik, Charité-Universitätsmedizin, Germany
Paper ID: 70

17:40-18:00 *Illumina 生物信息学整体解决方案*
唐顺江
Illumina 大中华区生物信息平台总监
Paper ID: 103

16:20-18:00 Session B3 (Tianzi 2 Conference Room at 1st floor 一楼天字二号会议室)

Topic: Systems Biology
Chair: Zengyou He

16:20-16:40 *A Petri nets-based framework for whole-cell modeling*
Fei Liu*, Hengjie Song
South China University of Technology, China
Paper ID: 11

16:40-17:00 *Identification subtypes of Non–Small Cell Lung Cancer*
Yuxuan Zhou, Fuyan Hu*
Wuhan University of Technology, China
Paper ID: 24

17:00-17:20 *MD-SVM: A novel SVM-based algorithm for the motif discovery of transcription factor binding sites*
Jialu Hu*, Jingru Wang, Xuequn Shang
Northwestern Polytechnical University, China
Paper ID: 30

17:20-17:40 *Detecting the stable point of therapeutic effect of chronic myeloid leukemia based on dynamic network biomarkers*
Junhua Xu[1], Min Wu[1], Shanshan Zhu[1], Jinzhi Lei[2], Jie Gao[1,*]
[1]Jiangnan University, China
[2]Zhou Pei-Yuan Center for Applied Mathematics, China
Paper ID: 32

17:40-18:00 *The Cis-trans Binding Strength Defined by Motif Frequencies Facilitates Statistical Inference of Transcriptional Regulation*
Von Weber, Lei Li*
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China

Paper ID: 38

18:30-20:00 Banquet

20:30-21:30 Board member meeting of CSBMB-MSB (Tianzi 2 Conference Room at 1st floor 一楼天字二号会议室)

August 20 (Monday) Technical sessions

08:30-10:30 Plenary Session P2 (Kendo Hall at 2nd floor 二楼剑道馆)

Chair: Zefeng Wang

08:30-09:10 *Towards a complete map of the human long non-coding RNA transcriptome*

Rory Johnson[1,2]

[1]Department of Biomedical Research, University of Berne, Bern, Switzerland

[2]Department of Medical Oncology, Inselspital, Bern University Hospital, University of Bern, Switzerland

09:10-09:50 *Computational models for studying single and collective cell dynamics*

Timothy Elston

University of North Carolina, USA

09:50-10:30 *Data-driven modelling of cellular network*

Mariko Okada-Hatakeyama[1,2]

[1]Laboratory of Cell Systems, Institute for Protein Research, Osaka University, Osaka, Japan

[2]RIKEN Center for Integrative Medical Sciences (IMS), Kanagawa, Japan

10:30-10:50 Coffee break

10:50-12:30 Highlight Session A4 (Kendo Hall at 2nd floor 二楼剑道馆)

Topic: Systems Biology

Chair: Anyuan Guo

10:50-11:10 *Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins*

Chuan Dong[1], Gefei Hao[2], Hongli Hua[1], Shuo Liu[1], Abraham Alemayehu Labena[1], Guoshi Chai[1], Jian Huang[1], Nini Rao[1], Fengbiao Guo[1,*]

[1]University of Science and Technology of China, China

[2]Central China Normal University, China

Paper ID: 18

11:10-11:30 *RISE: a database of RNA interactome from sequencing experiments*

Qiangfeng Zhang

Tsinghua University, China

Paper ID: 40

11:30-11:50 *The free energy cost of synchronization in caynobacterial post-transcriptional circadian clock*

Dongliang Zhang[1,*], Yuansheng Cao[2], Yuhai Tu[3] and Qi Ouyang[1]

[1]Peking University, China

[2]University of California San Diego, USA

[3]IBM W.J Research Center, USA
Paper ID: 46

11:50-12:10 *Pan-genome analyses of 24 Shewanella strains re-emphasize the diversification of their functions yet evolutionary dynamics of metal-reducing pathway*
Kang Ning[*] and Choafang Zhong
Huazhong University of Science and Technology, China
Paper ID: 72

12:10-12:30 *Building a Knowledgebase for Precision Medicine*
Lei Liu
Fudan University, China
Paper ID：83

10:50-12:30 Session B4 (Tianzi 2 Conference Room at 1st floor 一楼天字二号会议室)

Topic: Diseases and Drugs
Chair: Lei Li

10:50-11:10 *Predicting microRNA-disease associations based on microRNA structural and functional similarity network*
Tao Ding, Jie Gao[*], Shanshan Zhu, Junhua Xu, Min Wu
Jiangnan University, China
Paper ID: 16

11:10-11:30 *Multi-scale modeling reveals angiogenesis-induced drug resistance in brain tumor and predicts a synergistic drug combination targeting EGFR and VEGFR pathways*
Weishan Liang, Xiaoqiang Sun[*]
Sun Yat-Sen University, China
Paper ID: 20

11:30-11:50 *PWCDA: A New and Efficient Method for Predicting circRNA-disease Associations*
Xiujuan Lei[1,*], Zengqiang Fang[1], Luonan Chen[2], Fangxiang Wu[3]
[1]Shanxi Normal University, China
[2]Chinese Academy of Sciences, China
[3]University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada
Paper ID: 31

11:50-12:10 *ODAE: Ontology-based systematic representation and analysis of drug adverse events and its usage in study of adverse events given different patient age and disease conditions*
Hong Yu[1,2,*], Solomiya Nysak[3], Noemi Garg[3], Edison Ong[4], Yongqun He[5,*]
[1]Guizhou Provincial People's Hospital;
[2]Guizhou University Medical College, China
[3]College of Literature, Science, and the Arts, University of Michigan, United States
[4]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, United States
[5]University of Michigan, United States

Paper ID: 36

12:10-12:30 *Screening drug combinations in disease-related molecular network*

Min Luo, Jianfeng Jiao and Ruiqi Wang*

Department of Mathematics, Shanghai University, Shanghai, China

Paper ID: 61

12:30-13:30 Lunch

14:00-15:40 Highlight Session A5 (Kendo Hall at 2nd floor 二楼剑道馆)

Topic: Network Biology

Chair: Xianwen Ren

14:00-14:20 *Inference of differentiation time for single cell transcriptomes using cell population reference data*

Na Sun[1], Xiaoming Yu[2], Fang Li[1], Denghui Liu[1], Shengbao Suo[1], Weiyang Chen[1], Shirui Chen[3], Lu Song[3], Christopher D. Green[1], Joseph McDermott[1], Qin Shen[2], Naihe Jing[3] and Jing-Dong J. Han[1,*]

[1]Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, China

[2]Tsinghua-Peking Center for Life Sciences, Tsinghua University, China

[3]Shanghai Institutes for Biological Sciences, China

Paper ID: 81

14:20-14:40 *Cell Lysate Microarray for Mapping the Network of Genetic Regulators for Histone Marks*

Li Cheng[1], Jun-Biao Dai[2,*] and Sheng-Ce Tao[1,*]

[1]Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, China

[2]Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, CAS, China

Paper ID: 42

14:40-15:00 *Modeling Endoplasmic Reticulum Network Maintenance in a Plant Cell*

Congping Lin[1,*], Rhiannon White[2], Imogen Sparkes[3] and Peter Ashwin[2]

[1]Huazhong University of Science and Technology, China

[2]University of Exeter, UK

[3]University of Bristol, UK

Paper ID: 49

15:00-15:20 *Biosystems Study of the Molecular Networks Underlying Hippocampal Aging Progression and Anti-aging Treatment in Mice*

Jiao Wang[1], Qian Li[1], Yanyan Kong[2], Fangfang Zhou[1], Jie Li[1], Weihao Li[1], Kai Wang[3], Ting Wu[4], Yihui Guan[2], Jiang Xie[5,*] and Tieqiao Wen[1,*]

[1]Laboratory of molecular neural biology, School of life sciences, Shanghai University, China

[2]Position Emission Computed Tomography Center, Huashan Hospital, Fudan University, China

[3]Shanghai Key Laboratory of Molecular Andrology, Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, China
[4]Shanghai Stem Cell Group, China
[5]School of Computer Engineering and Science, Shanghai University, China
Paper ID: 65

15:20-15:40 *Deep learning helps optimize CRISPR guide RNA design*
Fangzhou Shen, Renliang Sun, Jian Li and Zhuo Wang*
Shanghai Jiao Tong University, China
Paper ID: 67

14:00-15:40 Session B5 (Tianzi 2 Conference Room at 1st floor 一楼天字二号会议室)

Topic: Diseases and Drugs
Chair: Jinzhi Lei

14:00-14:20 *Developing the system model to predict High-risk Intracerebral Hemorrhage Patients by using Computed Tomography Angiography data*
Le Zhang[1,*], Kaikai Yin[2], Zhouyang Jiang[3],Tingting Li[2], Jin Li[2], Rong Hu[3], Zheng Yu[3], Hua Feng[4], Yujie Chen[4,*]
[1]Sichuan University, China
[2]Southwest University, China
[3]Third Military Medical University, China
[4]Southwest Hospital, China
Paper ID: 7

14:20-14:40 *Multi-scale analysis of schizophrenia risk genes, brain structure, and clinical symptoms reveals integrative clues for subtyping schizophrenia patients*
Liang Ma[1], Edmund Rolls[2,8], Xiuqin Liu[3], Yuting Liu[4], Zeyu Jiao[5], Yue Wang[4], Weikang Gong[6], Jie Zhang[5], Zhiming Ma[7], Fuzhou Gong[7], Lin Wan[7,*], Jianfeng Feng[5,8,*]
[1]Beijing Institute of Genomics, CAS, China
[2]Oxford Centre for Computational Neuroscience, UK
[3]School of Mathematics and Physics, University of Science and Technology Beijing, China
[4]Beijing Jiaotong University, China
[5]School of Mathematical Sciences and Centre for Computational Systems Biology, Fudan University, China
[6]CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, CAS, China
[7]Academy of Mathematics and Systems Science, CAS, China
[8]Department of Computer Science, University of Warwick, UK
Paper ID: 14

14:40-15:00 *Diagnosing pre-disease state of influenza A disease based on dynamical network biomarkers*
Shanshan Zhu, Jie Gao*, Junhua Xu, Min Wu, Tao Ding
Jiangnan University, China

Paper ID: 33

15:00-15:20 *Biosystems study of the molecular networks underlying Alzheimer's disease: differences between wild-type and Atp11b-knockout mice*

Jiao Wang, Fangfang Zhou, Jie Li, Qian Li, Weihao Li, Fangfang Ma, Tieqiao Wen*
Laboratory of Molecular Neural Biology, School of Life Sciences, Shanghai University, China

Paper ID: 39

15:20-15:40 *A novel joint gene set analysis framework improves identification of enriched pathways in cross disease transcriptomic analysis*

Wenyi Qin[1], Xujun Wang[2] and Hui Lu[3,*]
[1]Yale University, USA
[2]Shanghai Jiao Tong University, China
[3]University of Illinois at Chicago, USA

Paper ID: 60

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

16:20-18:00 Highlight Session A6 (Kendo Hall at 2nd floor 二楼剑道馆)

Topic: Systems Biology

Chair: Yong Wang

16:20-16:40 *Quantifying the biological functions in gene regulatory networks*

Lei Zhang
Peking University, China

Paper ID: 4

16:40-17:00 *Parameter sensitivity analysis for a stochastic model of mitochondrial apoptosis pathway*

Xianli Chen[1,*], Xiaoguang Li[2], Wei Zhao[3], Qi Ouyang[4] and Tiejun Li[5]
[1]Peking University, China
[2]College of Mathematics and Compute Science, Hunan Normal University, Changsha, China
[3]Center for Quantitative Biology and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China
[4]The State Key Laboratory for Artificial Microstructures and Mesoscopic Physics, Department of Physics, China
[5]LMAM and School of Mathematical Sciences, Peking University, Beijing, China

Paper ID: 44

17:00-17:20 *Cancer development as a dynamical system*

Jinzhi Lei
Tsinghua University, China

Paper ID: 45

17:20-17:40 *Systems Biology Theory Resolution of a Controversy in Pancreatic Beta Cell Regeneration*

Haoran Cai[1], Runtan Cheng[2], Xiaomei Zhu[1] and Ping Ao[1,*]
[1]Shanghai University, China
[2]Shanghai Jiao Tong University, China
Paper ID: 75

17:40-18:00 *Integrated regulatory-metabolic network modeling and strain design based on the integrated model*

Guohui Chuai[1,2], Hanhui Ma[5], Jifang Yan[1,2], Ming Chen[4], Nanfang Hong[1,2], Dongyu Xue[1,2], Chi Zhou[1,2], Chenyu Zhu[1,2], Ke Chen[1,2], Bin Duan[1,2], Feng Gu[6], Sheng Qu[1,2], Deshuang Huang[3,*], Jia Wei [4,*]and Qi Liu[1,2,*]
[1]Department of Endocrinology & Metabolism, Shanghai Tenth People's Hospital, Tongji University, Shanghai 20009, China
[2]Bioinformatics Department, School of Life Science and Technology, Tongji University, Shanghai 20009, China
[3]Machine learning & Systems Biology Lab, School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China
[4]R&D Information, Innovation Center China, AstraZeneca, 199 Liangjing Road, Shanghai 201203, China
[5]School of Life Science and Technology, ShanghaiTech University, Shanghai, China
[6]State Key Laboratory Cultivation Base and Key Laboratory of Vision Science, Ministry of Health and Zhejiang Provincial Key Laboratory of Ophthalmology and Optometry, School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China
Paper ID: 105

16:20-18:00 Session B6 (Tianzi 2 Conference Room at 1st floor 一楼天字二号会议室)

Topic: Proteomics

Chair: Binqiang Liu

16:20-16:40 *PhD7Faster 2.0: predicting clones propagating faster from the Ph.D.-7 phage display library by coupling PseAAC and tripeptide composition*

Bifang He*, Ning Li, Jian Huang*
University of Science and Technology of China, China
Paper ID: 13

16:40-17:00 *PTPD: Prediction of Therapeutic Peptides by Deep Learning*

Chuanyan Wu, Rui Gao*, Yusen Zhang
Shandong university, China
Paper ID: 26

17:00-17:20 *A voting mechanism-based linear epitope prediction system for the host-specific Iridoviridae family*

Tunwen Pai
Department of Computer Science and Engineering, National Taiwan Ocean University, Taiwan
Paper ID: 37

17:20-17:40 *Computational prediction and functional analysis of arsenic binding proteins in human cells*

Shichao Pang[1], Junchen Yang[1], Yilei Zhao[1], Yixue Li[2] and Jingfang Wang[1,*]

[1]Shanghai Jiao Tong University, China

[2]Chinese Academy of Sciences, China

Paper ID: 59

17:40-18:00 *DOS: A tool for predicting degree of specificity of monoclonal antibodies using sequences*

Anthony Mackitz Dzisoo and Jian Huang[*]

University of Electronic Science and Technology of China, Chengdu, China.

Paper ID: 62

18:30-20:00 Dinner

August 21 (Tuesday)

08:00-13:00 Half day excursion in Qingyan Ancient Town. Departure at 8:00 from lobby

Poster Session

Robust Feature Extraction by Sample Pattern Embedding of single-cell RNA-seq data
Qianqian Shi[1], Chuanchao Zhang[2,*], Luonan Chen[3,*]
[1]College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
[2]Shanghai Institutes for Biological Sciences, China
[3]Osaka Sangyo University, China
Paper ID: 15

MicroRNAs in tsetse fly (Glossina morsitans) reveal novel insight to sleeping sickness
Zhiyuan Yang[1,*], Xi Zeng[2], Angel Tsz-Yau Wan[2]
[1]Hangzhou Dianzi University, China
[2]The Chinese University of Hong Kong, Hong Kong
Paper ID: 19

Network Analysis of Single-cell RNA Sequencing Data based on cell-specific network
Hao Dai, Lin Li, Tao Zeng and Luonan Chen[*]
Shanghai Institute of Biochemistry and Cell Biology, CAS, China
Paper ID: 48

Sequential data analysis based on constrained optimization by bivariate PCA
Si Zhang, Tao Zeng and Luonan Chen[*]
Institute of Biochemistry and Cell Biology, China
Paper ID: 50

Predicting Conserved Regions in Protein Sequences Using Equivalence Classes
Jingsong Zhang[*], Tao Zeng and Luonan Chen
Institute of Biochemistry and Cell Biology, CAS, China
Paper ID: 51

Network clustering of Single-cell RNA Sequencing Data
Lin Li and Luonan Chen[*]
CAS, China
Paper ID: 52

Analysis of Hyperdysregulatory network in Human with Pancancer Data
Pingyang Wang[1,*], Lina Lu[1,*], Tao Zeng[2,*] and Luonan Chen[2,*]
[1]Shanghai Institute of Biochemistry and Cell Biology, China

[2]Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, CAS, China

Paper ID: 53

Analysis of molecular mechanisms for EGFR-TKI resistance in NSCLC based on transcriptome data

Tang Shijie and Chen Luonan[*]

Shanghai Institute of Biochemistry and Cell Biology, China

Paper ID: 54

Revealing the Tipping Points in Infant Brain Development for Human and Chimpanzee by Gene Expression Data

Hui Tang, Ying Tang, Tao Zeng and Luonan Chen[*]

Chinese Academy of Sciences, Shanghai 200031, China

Paper ID: 63

Personalized critical variation of gut microbiota before the occurrence of Type I Diabetes

Lu Wang[1,*] and Luonan Chen[2,*]

[1]SIBCB, China

[2]Shanghai Institutes for Biological Sciences, China

Paper ID: 64

The carcinogenic molecules were revealed by novel integrative analysis

Wanting Liu[*] and Gong Zhang[*]

Jinan University, China

Paper ID: 73

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

Jing Ge[1], Gaopeng Li[2], Shuxian Li[3], Huiyong Yin[4,*] and Luonan Chen[5,*]

[1]Institute of Biochemistry and Cell Biology, SIBS, CAS, China

[2]Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China

[3]School of Life Science and Technology, ShanghaiTech University, China

[4]Institute for Nutritional Sciences, SIBS, CAS, China

[5]Osaka Sangyo University, China

Paper ID: 74

Systematic survey and prediction reveal widespread context-dependent activities of RNA binding proteins in splicing regulation

Yue Hu, Miaowei Mao and Zefeng Wang[*]

CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, China

Paper ID: 77

RBM10 functions as a tumor suppressor in lung cancer by mediating alternative splicing of key target genes

Sirui Zhang[1], Yongbo Wang[2,*] and Zefeng Wang[1,*]

[1]CAS-MPG Partner Institute for Computational Biology, China

[2]Fudan university, China

Paper ID: 79

Subnetwork identification and chemical modulation for neural regeneration. A study combining network guided forest and heat diffusion model

Hui Wang[1], Gang Wang[1], Lida Zhu[1], Xuan Xu[1], Bo Diao[2] and Hong-Yu Zhang[1,*]

[1]Huazhong Agricultural University, China

[2]Department of Clinical Experiment, Wuhan General Hospital of Guangzhou Command, China

Paper ID: 80

Identifying the patterns of double-stranded break sites during meiosis homologous recombination

Qiu Wang

Shanghai Institute of Biochemistry and Cell Biology, CAS, China

Paper ID: 82

Distinct endophytic communities associate with different plants for adaptation to karst environments

Fei Li, Xiaohong He, Yuanyuan Sun, Ximin Zhang, Xiaoxin Tang, Yuke Li, Yin Yi*

The Key Laboratory of biodiversity conservation in Karst mountain area of Southwest of china, Forestry Ministry, Guizhou Normal University

Key Laboratory of Plant Physiology and Developmental Regulation, Guizhou Normal University

School of Life Sciences, Guizhou Normal University, Guiyang, Guizhou, China

Paper ID: 104

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

Book of Abstracts

Plenary Sessions

Deciphering developmental patterning with deep neural network

Jingxiang Shen[1], Feng Liu[1,2],Chao Tang[1,2,3]

[1]Center for Quantitative Biology, Peking University, Beijing 100871, China

[2]School of Physics, Peking University, Beijing 100871, China

[3]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

Dynamics of complex biological systems is driven by intricate networks, the current knowledge of which are often incomplete. The traditional systems biology modeling usually implements an ad hoc fixed set of differential equations with predefined function forms. Such an approach often suffers from over-fitting the data and inadequate predictive power, especially when dealing with systems of high complexity. This problem could be overcome by deep neuron network. Choosing pattern formation of the gap genes in Drosophila early embryogenesis as an example, we established a deep neural network (DNN). The trained DNN model yields perfect fitting and impressively accurate predictions on mutant patterns. We further mapped the trained DNN into a simplified conventional regulation network, which is consistent with the existing knowledge. The DNN model could lay a foundation of "in-silico-embryo" on which one can perform all kinds of perturbations to discover underlying mechanisms. This approach can be readily applied to a variety of complex biological networks.

Brief CV

Chao Tang is a Chair Professor of Physics and Systems Biology at Peking University. He had his undergraduate training at the University of Science and Technology of China and received a Ph.D. degree in Physics from the University of Chicago. In his early career, he worked on problems in statistical physics, condensed matter physics, dynamical and complex systems. His current research interest is at the interface between physics and biology, in particular in quantitative systems biology and biological physics. He was a tenured full professor at the University of California San Francisco before returning to China full-time in 2011. He is a Fellow of the American Physical Society, the founding director of the interdisciplinary Center for Quantitative Biology at Peking University and the founding Co-Editor-in-Chief of the journal Quantitative Biology.

Machine learning and complex networks for complex systems big data analysis and precision medicine

Carlo Vittorio Cannistraci
Biomedical Cybernetics Group, Technical University Dresden, Germany

The talk will present our research at the Biomedical Cybernetics Group that I established about four years ago in Technical University Dresden. We adopt a transdisciplinary approach integrating information theory, machine learning and network science to investigate the physics of networked adaptive complex systems at different scales, from molecules to ecological and social systems, with a particular attention to biology and medicine, and a new emerging interest for the analysis of complex big data in social and economic science. Our theoretical effort is to translate advanced mathematical paradigms typically adopted in theoretical physics (such as topology, network and manifold theory) to characterize many-body interactions in complex systems. We apply the theoretical frameworks we invent in the mission to develop computational tools for systems and network analysis. In particular, we deal with: prediction of wiring in networks and multiscale-combinatorial marker design for quantification of topological modifications in complex networks. Our attention for precision biomedicine is aimed to topics with important impact from the economical point of view such as development of tools for disease biomarker discovery, drug repositioning and combinatorial drug therapy.
This talk will focus on two main theoretical innovation. Firstly, the development of machine learning for topological estimation of nonlinear relations in high-dimensional data[1] (or in complex networks[2]) and its relevance for applications in big data, with a particular emphasis on biomedicine. Secondly, we will discuss the Local Community Paradigm (LCP)[4,5], which is a theory proposed to model local-topology-dependent link-growth in complex networks and therefore it is useful to devise topological methods for link prediction in monopartite and bipartite[5] networks such as molecular drug-target interactions[6] and product-consumer networks.

References

1.Cannistraci, C. V., Ravasi, T., Montevecchi, F. M., Ideker, T. & Alessio, M. Nonlinear dimension reduction and clustering by minimum curvilinearity unfold neuropathic pain and tissue embryological classes. Bioinformatics 26, i531–i539 (2010).
2.Cannistraci, C. V., Alanis-Lobato, G. & Ravasi, T. Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. in Bioinformatics 29, (2013).
3.Muscoloni, A., Thomas, J. M., ..., & Cannistraci, C. V. Machine learning meets complex networks via coalescent embedding of networks in the hyperbolic space. Nature Communication (2017).
4.Cannistraci, C. V., Alanis-Lobato, G. & Ravasi, T. From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks. Sci. Rep. 3, 1–13 (2013).
5.Daminelli, S., Thomas, J. M., Durán, C. & Vittorio Cannistraci, C. Common neighbours and the local-community-paradigm for topological link prediction in bipartite networks. New J. Phys. 17, 113037 (2015).
6. Duran, C.,…, Cannistraci, C.V. Pioneering topological methods for network-based drug–target prediction by exploiting a brain-network self-organization theory. Briefings in Bioinformatics,

bbx041 (2017).
7.Ravasi, T.*, Cannistraci C.V.*, et al. An atlas of combinatorial transcriptional regulation in mouse and man. Cell 140, (2010).
8.Ammirati, E.*, Cannistraci, C.V.*, et al. Identification and predictive value of interleukin-6+ interleukin-10+ and interleukin-6-interleukin-10+ cytokine patterns in st-elevation acute myocardial infarction. Circ. Res. 111, 1336–1348 (2012).
9.Narula, V., …, and Cannistraci, C.V. Can local-community-paradigm and epitopological learning enhance our understanding of how local brain connectivity is able to process, learn and memorize chronic pain? Applied Network Science 2 (1), 28
10.Cannistraci, C.V., et al., "Summer Shift": A Potential Effect of Sunshine on the Time Onset of ST-Elevation Acute Myocardial Infarction. Journal of the American Heart Association 7 (8), e006878

Brief CV

Carlo Vittorio Cannistraci is a theoretical engineer, head of the Biomedical Cybernetics Group and faculty of the Department of Physics in the Technical University Dresden, which is a member of the TU9 excellence-league (the nine most prestigious technical universities in Germany). Carlo's area of research embraces information theory, machine learning and complex networks including also applications in systems biomedicine and neuroscience. Nature Biotechnology selected Carlo's article (Cell 2010)7 on machine learning in developmental biology to be nominated in the list of 2010 notable breakthroughs in computational biology. Circulation Research featured Carlo's work (Circulation Research 2012)[8] on leveraging a cardiovascular systems biology strategy to predict future outcomes in heart attacks, commenting: "a space-aged evaluation using computational biology". The Technical University Dresden honoured Carlo of the Young Investigator Award 2016 in Physics for his work on the local-community-paradigm theory and link prediction in bipartite networks[5]. In 2017, Springer-Nature scientific blog highlighted with an interview to Carlo his study on "How the brain handles pain through the lens of network science"[9]. The American Heart Association covered this year on its website the recent chronobiology discovery of Carlo on how the sunshine affects the risk and time onset of heart attack[10].

Towards a complete map of the human long non-coding RNA transcriptome

Rory Johnson[1,2]

[1]*Department of Medical Oncology, Inselspital, Bern University Hospital, University of Bern, Switzerland.*

[2]*Department for BioMedical Research, University of Bern, Bern, Switzerland*

Accurate annotations of genes and their transcripts is a foundation of genomics, but no annotation technique presently combines throughput and accuracy. As a result, current reference gene collections remain far from complete: many genes models are fragmentary, while thousands more remain uncatalogued—particularly for long noncoding RNAs (lncRNAs). To accelerate lncRNA annotation, the GENCODE consortium has developed RNA Capture Long Seq (CLS), combining targeted RNA capture with third generation long-read sequencing. CLS outperforms short-read assemblies in the depth and completeness of its annotations, yielding full-length lncRNA transcript models for the first time. I will discuss how GENCODE plans to develop CLS to eventually achieve complete annotation of the non-coding transcriptome in human and mouse. Time permitting, I will also present examples of how accurate lncRNA maps can be employed to hunt for new cancer genes.

Brief CV

Rory Johnson's research aims to understand the role of long non-coding RNAs in disease, using a combination of bioinformatic and experimental tools. In 2007 he obtained his PhD, funded by the Wellcome Trust at the University of Leeds, applying genomic microarray and bioinformatic motif analysis to the regulation of microRNAs in neurodegeneration. Next, a postdoctoral position at the Genome Institute of Singapore exposed him to next generation sequencing and kindled my interest in little-known genes called long non-coding RNAs. This led him to the position of Staff Scientist with Roderic Guigo at the CRG, Barcelona, where he began work with GENCODE on lncRNA annotations. In 2016 he established the Genomics of Long Noncoding RNA in Disease (GOLD) Lab within the Medical Faculty at the University of Bern, funded by the NCCR "RNA and Disease" project. Currently the main interests are how to complete the human lncRNA annotation, and how to use CRISPR-Cas9 to find clinically-actionable lncRNAs in cancer and heart disease.

Computational models for studying single and collective cell dynamics

Timothy Elston
University of North Carolina, USA

Almost all cells are able to grow or move in a directed manner. The first step in generating directed movement is polarity establishment. That is, the establishment of a cell front and back. We combine computational and experimental approaches to understand the mechanisms that underlie polarity establishment at the single cell level and the collective behavior that emerges when multiple cells interact. In particular, we use stochastic models to understand polarity establishment in yeast and the mechanisms that ensure only a single cell front is produced. Next we present a model for the collective behavior of multiple endothelial cells. We use the model to investigate the role of cerebral cavernous malformation (CCM) proteins in the early stages of vascular tube formation.

Brief CV

Timothy Elston received his graduate training in physics with an emphasis on statistical physics and nonlinear dynamics. As a postdoctoral researcher, he became interested in applying tools from these fields to problems in biophysics and cell biology. Currently, his lab integrates computational approaches, including mathematical modeling and quantitative image analysis, with experimental investigations to understand complex cellular behavior. His lab is particularly interested in understanding the molecular mechanisms that regulate the spatiotemporal dynamics of cell signaling networks. They also develop novel computational techniques for quantitative analyses of live-cell images and simulating spatiotemporal models of signaling pathways. Current projects in the lab focus on cell fate decisions, polarity establishment and gradient sensing. The primary model system they use to study these cellular functions is the yeast Saccharomyces cerevisiae. Their investigations combine microfluidic technology with live-cell microscopy to observe cellular behavior in well-controlled environments. This experimental platform provides a powerful system for developing and validating predictive models of cellular function. The lab also is involved in multiple collaborative projects to investigate these fundamental cellular processes in physiological contexts and their dysregulation in human disease.

Throughout his career he has made significant contributions to the research community both within and outside the University of North Carolina at Chapel Hill (UNC-CH). At UNC-CH, he is currently the Director of the Ph.D. Curriculum in Bioinformatics and Computational Biology and Co-Director of the recently created Computational Medicine Program at UNC. He also serves as Chair of the Research Computing Advisory Committee. Outside of the university, he has served as a standing member of the NIH Modeling and Analysis of Biological Systems (MABS) Study Section and has served on multiple NSF review panels. He is on the Board of Reviewing Editors for Science Magazine and the editorial board of SIAM Dynamical Systems. Currently, he serves on the Advisory Board for the San Diego Center for Systems Biology at UCSD, and he is a past member of the advisory board for the Mathematical Biosciences Institute at the Ohio State University.

Data-driven modelling of cellular network

Mariko Okada-Hatakeyama[1,2]

[1]*Laboratory of Cell Systems, Institute for Protein Research, Osaka University, Osaka, Japan*

[2]*RIKEN Center for Integrative Medical Sciences (IMS), Kanagawa, Japan*

Signal transduction pathways transmit an extracellular information to nuclear transcription factors to activate gene expression for cell fate decision. Interestingly, the pathways often control this process in a highly nonlinear manner, and in some cases, an analogous change in extracellular information can be transformed into a digital form of transcription factor activity. Digital all-or-none activity of the transcription factors seems to be important to achieve accurate cellular processing in noisy intracellular environment. Such examples include c-Fos and NF-kappa B transcription factors activated by ERK and IKK signaling activities, respectively.

Next question is how these digitally activated transcription factors control gene expression. To understand a quantitative relationship between the transcription factor activity and downstream target gene expression, we obtained several types of NGS data for mRNA, histone modification and chromatin accessibility and analyzed the data using bioinformatics tools and kinetic models. Our analysis indicates that a positive cooperativity in epigenetic regulation control all-or-none activation of target gene expression, and this mechanism might act as a threshold mechanism for cell fate decision. I will introduce our approach for NF-□B signaling pathway in immune B cells.

References

Inoue et al. (2016) npj Syst. Biol. Appl. 2, 16024.

Shinohara et al. (2014) Science. 344(6185):760-764.

Nakakuki et al. (2010) Cell. 141(5):884-896.

Brief CV

Name Mariko Okada (Hatakeyama), Ph.D.

E-mail mokada@protein.osaka-u.ac.jp

EDUCATION/TRANING

Ph.D., Tokyo University of Agriculture and Technology

Doctoral research at Institute of Toxicology and Environmental Health, University of California, Davis, USA

ACADEMIC POSITIONS

2016- Professor, Institute of Protein Research, Osaka University

2013- Team Leader, RIKEN Center for Integrative Medical Sciences (IMS)

2009- 2013 Team leader, RIKEN Center for Allergy and Immunology (RCAI)

2008- 2009 Team leader, RIKEN Advanced Science Institute (ASI)

2005- 2008 Team leader, RIKEN Genome Sciences Center (GSC)

2000- 2005 Research scientist, RIKEN Genome Sciences Center (GSC)

RESEARCH INTERESTS

Mathematical modeling of signaling network

Time-development of cellular network

Cellular specificity by molecular cooperativity

Parallel and Poster Sessions

Paper ID: 3

CHDOCK: A hierarchical docking approach for predicting the structure of homo-oligomeric complexes with Cn symmetry

*Yu-Meng Yan and Sheng-you Huang**

Huazhong University of Science and Technology, China

Protein-protein interactions are crucial in many biological processes. Therefore, determining the complex structure between proteins is valuable for understanding the molecular mechanism and developing drugs. Many proteins like ion channels are formed by symmetric homo-oligomers. In this study, we have proposed a hierarchical docking algorithm to predict the structure of Cn symmetric protein complexes, which is referred to as CHDOCK. The symmetric binding modes were first constructed by an FFT-based docking algorithm and then optimized through our iterative scoring function for protein-protein interactions. When tested on a symmetric protein docking benchmark of 212 homo-oligomeric complexes with Cn symmetry, CHDOCK obtained a significantly better performance in binding mode predictions than three state-of-the-art symmetric docking methods, M-ZDOCK, SAM, and SymmDock. When the top 10 predictions were considered, CHDOCK achieved a success rate of 44.81% and 72.17% for idealized bound docking and realistic unbound docking, compared to 36.79% and 65.09% for M-ZDOCK, 31.60% and 54.24% for SAM, and 30.66% and 31.60% for SymmDock, respectively. CHDOCK is computationally efficient and can normally complete a symmetric docking calculation within 30 minutes. The CHDOCK can be freely accessed through a web server at http://huanglab.phys.hust.edu.cn/hsymdock/.

Paper ID: 4

Quantifying the biological functions in gene regulatory networks

Lei Zhang

Peking University, China

Gene regulatory network in biology plays a critical role in achieving accurate biological functions. In this talk, I will start with exploration of the topologies for dual function networks by achieving both adaptation and noise attenuation. We show the three-node networks are not able to buffer noise while implementing a good adaptation due to the tradeoff. Thus, we construct a four-node network topology achieves dual functions, in which the fluctuation in input is dampen significantly in the upstream reactions and the downstream reactions accomplish the adaptive behavior. Secondly, I will present the dual role of Nanog during stem cell differentiation and reprogramming. The low-Nanog state enhances cell differentiation through serving as an intermediate state to reduce the energy barrier of transition. On the contrary, the existence of low stemness low-Nanog state will slow down the reprogramming process, and additional Nanog activation is revealed to be essential to attain fully reprogrammed cell state faster.

Paper ID: 6

CrystalM: a multi-view fusion method for protein crystallization prediction

Yubo Wang[1], Jijun Tang[2], Yu Dai[1], Fei Guo[1*]*

[1]Tianjin University, China

[2]University of South Carolina, United States

Improving the accuracy of predicting protein crystallization is very important for protein crystallization projects, which is a critical step for the determination of protein structure by X-ray crystallography. At present, many methods including machine learning methods are used to predict protein crystallization. Here, we use a novel feature combination to construct SVM model in the prediction of protein crystallization, called as CrystalM. In this work, we extract six features to represent protein sequences, namely Average Block-Position specific scoring matrix (AVBlock-PSSM), Average Block-Secondary Structure (AVBlock-SS), Global Encoding (GE), Pseudo-Position specific scoring matrix (PsePSSM), Protscale and Discrete Wavelet Transform-Position specific scoring matrix (DWT-PSSM). Moreover, we employed two training datasets (TRAIN3587 and TRAIN1500) and their corresponding independent test datasets (TEST3585 and TEST500) to evaluate our method by feeding multi-view features into Support Vector Machine (SVM) classifier. Two training datasets are employed for five-fold cross validation, and two test datasets are separately used to test the corresponding datasets. Finally, we compared our method with other existing methods in the performance. For the datasets of TRAIN3587 and TEST3585, our method achieved best Accuracy (ACC), best Specificity (SP) and the same Mathew's correlation coefficient (MCC) as the previous outperforming method in the five-fold cross validation. In particular, ACC, SP and MCC have surpassed the existing methods in independent test, which proves the effectiveness of our method. Meanwhile, ACC, SP and MCC are higher than existing methods in the five-fold cross validation for TRAIN1500. Although the performance of independent test for TEST500 is not the best, our method also has a certain predictability in the prediction of protein crystallization. In addition, we find that only choosing the first four features can improve the performance of prediction for TRAIN1500 and TEST500, not only in independent tests but also in five-fold cross validation. This phenomenon indicates that the latter two features can not effectively represent proteins of TRAIN1500 and TEST500. Our method is a sequence-based protein crystallization prediction method. The good performance on the datasets proves the effectiveness of our method and the better performance on large datasets further demonstrates the stability and superiority of our method.

Paper ID: 7

Developing the system model to predict High-risk Intracerebral Hemorrhage Patients by using Computed Tomography Angiography data

Le Zhang[1], Kaikai Yin[2], Zhouyang Jiang[3], Tingting Li[2], Jin Li[2], Rong Hu[3], Zheng Yu[3], Hua Feng[4] ,Yujie Chen[4*]*

[1]Sichuan University, China

[2]Southwest University, China

[3]Third Military Medical University, China

[4]Southwest Hospital, China

Background and Purpose: Though it is well known that Hemorrhagic stroke accounts for about 31.52% of all strokes and hypertension is the common origin, little is known about the way to identify the high-risk population of hypertensive intracerebral hemorrhage patients.
Methods: We employ experimental design, statistical test and classification algorithms to develop a predictive model for intracerebral hemorrhage with ten potential computed tomography angiography features around the basal ganglia and thalamus area, because previous research report that the most frequent occurrence location of hypertensive intracerebral hemorrhage is around this area.
Results: The results turn out that the angle between the middle cerebral artery and the internal carotid artery (AMIC), the distance between the beginning of median artery and superior trunk (DMS), the CT of lenticulostriate arterial (CTL) are statistically significant enough to be reason for intracerebral hemorrhage. In addition, we choose these three potential features as the classifiers for the ensemble learning classification model. Our developed ensemble-learning method outperforms than not only the previous work, but also other three classical classification methods in accuracy measurement.
Conclusions: The developed system model is efficient for the prediction of the incidence probability of intracerebral hemorrhage.

Paper ID: 8

Data Science to Invent Digital Ag Solution

Le Lv[1], and Jingdong Liu[2]*
[1]Monsanto Biotech(Beijing) Research Center, China
[2]Monsanto Company, United States

Digital agriculture has developed very fast in the last several years. Through working with The Climate Corporation, Monsanto use data science to increase efficiency by leveraging data about weather, soil and other patterns to help growers use agriculture solutions at the right time, in the right place. 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. Here we show some examples on how to collect and utilize different source of data to obtain deeper understanding of crop traits.

Paper ID: 9

Computational Design of Antiangiogenic Peptibody by Fusing Human IgG1 Fc Fragment and HRH Peptide: Structural Modeling, Energetic Analysis, and Dynamics Simulation of Its Binding Potency to VEGF Receptor

Lin Ning, Zhongyan Li, Zhengya Bai, Shasha Hou, Bifang He, Jian Huang, Peng Zhou**
Center for Information Biology, University of Electronic Science and technology of China, China

Peptibodies represent a new class of biological therapeutics with combination of peptide activity and antibody-like properties. Previously, we discovered a novel peptide HRH that exhibited a dose-

dependent angiogenesis-suppressing effect by targeting vascular endothelial growth factor receptors (VEGFRs). Here, we computationally designed an antiangiogenic peptibody, termed as PbHRH, by fusing the HRH peptide to human IgG1 Fc fragment using the first approved peptibody drug Romiplostim as template. The biologically active peptide of Romiplostim is similar with HRH peptide; both of them have close sequence lengths and can fold into a α-helical conformation in free state. Molecular dynamics simulations revealed that the HRH functional domain is highly flexible, which is functionally independent of Fc fragment in the designed PbHRH peptibody. Subsequently, the intermolecular interactions between VEGFR-1 domain 2 (D2) and PbHRH were predicted, clustered and refined into three representatives. Conformational analysis and energetic evaluation unraveled that the PbHRH can adopt multiple binding modes to block the native VEGF-A binding site of VEGFR-1 D2 with its HRH functional domain, although the binding effectiveness of HRH segments in peptibody context seems to be moderately decreased relative to that of free HRH peptide. Overall, it is suggested that integrating HRH peptide into PbHRH peptibody does not promote the direct intermolecular interaction between VEGFR-1 D2 and HRH. Instead, the peptibody may indirectly help to improve the pharmacokinetic profile and bioavailability of HRH.

Paper ID: 10

GSCALite:A Web Server for Gene Set Cancer Analysis

Anyuan Guo
Huazhong University of Science and Technology, China

The availability of cancer genomic data makes it possible to analyze genes related to cancer. Cancer is usually the result of a set of genes and the signal of a single gene could be covered by background noise. Here, we present a web server named Gene Set Cancer Analysis (GSCALite) to analyze a set of genes in cancers with the following functional modules. (i) Differential expression in tumor versus normal, and the survival analysis; (ii) Genomic variations and their survival analysis; (iii) Gene expression associated cancer pathway activity; (iv) miRNA regulatory network for genes; (v) Drug sensitivity for genes; (vi) Normal tissue expression and eQTL for genes. GSCALite is a user-friendly web server for dynamic analysis and visualization of gene set in cancer and drug sensitivity correlation, which will be of broad utilities to cancer researchers.

Paper ID: 11

A Petri nets-based framework for whole-cell modeling

Fei Liu[*]*, Hengjie Song*
South China University of Technology, China

Whole-cell models have been recognized as the central aim of systems biology but also a grant challenge, which play essential roles in current and future systems biology. In this paper, we comprehensively analyze whole-cell modeling requirements and classify them into three aspects (or dimensions): heterogeneous biochemical networks, uncertainties in components and representation of cell structure. We then explore how to use different Petri net classes to address different aspects of whole-cell modeling requirements. Based on these analyses we present a Petri nets-based

framework for whole-cell modeling. We hope this framework can offer a feasible modeling approach for whole-cell model construction, and address many of their challenges.

Paper ID: 12

Cross-species Data Classification by Domain Adaptation via Discriminative Heterogeneous Maximum Mean Discrepancy

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

Cross-species or Cross-platform data classification is a challenging problem in the field of Bioinformatics, which aims to classify data samples in one species/platform by using labelled data samples in another species/platform. Traditional classification methods can not be used in this case, since the samples from two species/platform may have different feature spaces, or follow different statistical distributions. Domain adaptation is a new strategy which could be used to deal with this problem. A big challenge in domain adaptation is how to reduce the difference and correct the drift between the source and the target domains in the heterogeneous case, when the feature spaces of the two domains are different. It has been shown theoretically that probability divergences between the two domains such as maximum mean discrepancy (MMD) play an important role in the generalization bound for domain adaptation. However,
they are rarely used for heterogeneous domain adaptation due to the different feature spaces of the domains. In this work, we propose a heterogeneous domain adaptation approach by making use of MMD, which measures the probability divergence in an embedded low-dimensional common subspace. Our proposed discriminative heterogeneous MMD approach (DMMD) aims to find new representations of the samples in a common subspace by minimizing the domain probability divergence with preserving the known discriminative information. A conjugate gradient algorithm on a Grassmann manifold is applied to solve the nonlinear DMMD model. Our experiments on both simulation and benchmark machine learning datasets show that our approaches outperform other state-of-art approaches for heterogeneous domain adaptation. We finally apply our approach to a cross-platform dataset and a cross-species dataset, and the results show the effectiveness of our approach.

Paper ID: 13

PhD7Faster 2.0: predicting clones propagating faster from the Ph.D.-7 phage display library by coupling PseAAC and tripeptide composition

Bifang He[*], Ning Li, Jian Huang[*]
University of Science and Technology of China, China

Selection from phage display libraries empowers isolation of high-affinity ligands for various targets. However, this method also identifies propagation-related target-unrelated peptides (PrTUPs). These false positive hits appear because of their amplification advantages. In this report, we present PhD7Faster 2.0 for predicting fast-propagating peptides from the Ph.D.-7 phage display library, which was developed based on support vector machine (SVM). Feature selection was performed

against PseAAC and tripeptide composition using the fselect method and the incremental feature selection method. Ten-fold cross-validation results show that PhD7Faster 2.0 succeeds a decent performance with the accuracy of 81.84%, the Matthews correlation coefficient (MCC) of 0.64 and the area under the ROC curve (AUC) 0.90. The permutation test with 1000 shuffles resulted in $p < 0.001$. We implemented PhD7Faster 2.0 into a publicly accessible web tool (http://i.uestc.edu.cn/sarotup3/cgi-bin/PhD7Faster.pl) and constructed standalone graphical user interface (GUI) and command-line versions for different systems. The standalone PhD7Faster 2.0 is able to detect PrTUPs within small datasets as well as large-scale datasets. This makes PhD7Faster 2.0 an enhanced and powerful tool for scanning and reporting faster-growing peptides from the Ph.D.-7 phage display library.

Paper ID: 14

Multi-scale analysis of schizophrenia risk genes, brain structure, and clinical symptoms reveals integrative clues for subtyping schizophrenia patients

Liang Ma[1], Edmund Rolls[2,8], Xiuqin Liu[3],Yuting Liu[4], Zeyu Jiao[5], Yue Wang[4], Weikang Gong[6], Jie Zhang[5], Zhiming Ma[7], Fuzhou Gong[7], Lin Wan[7], Jianfeng Feng[5,8*]*

[1]Beijing Institute of Genomics, CAS, China

[2]Oxford Centre for Computational Neuroscience, UK

[3]School of Mathematics and Physics, University of Science and Technology Beijing, China

[4]Beijing Jiaotong University, China

[5]School of Mathematical Sciences and Centre for Computational Systems Biology, Fudan University, China

[6]CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, CAS, China

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

[8]Department of Computer Science, University of Warwick, UK

Analysis linking directly genomics, neuroimaging phenotypes and clinical measurements is crucial for understanding psychiatric disorders, but remains rare. Here, we describe a multi-scale analysis using whole-genome SNPs, gene-expression, grey matter volume (GMV) and the Positive and Negative Syndrome Scale scores (PANSS) to explore the etiology of schizophrenia. With 72 drug-naive schizophrenic first episode patients (FEPs) and 73 matched heathy controls, we identified 108 genes, from schizophrenia risk genes, that correlated significantly with GMV, which are highly co-expressed in the brain during development. Among these 108 candidates, 19 distinct genes were found associated with 16 brain regions referred to as hot clusters (HCs), primarily in the frontal cortex, sensory-motor regions and temporal and parietal regions. The patients were subtyped into three groups with distinguishable PANSS scores by the GMV of the identified HCs. Furthermore we found that HCs with common GMV among patient groups are related to genes that mostly mapped to pathways relevant to neural signaling, which are associated with the risk for schizophrenia. Our results provide an integrated view of how genetic variants may affect brain structures that lead to distinct disease phenotypes. The method of multi-scale analysis that was described in this research, may help to advance the understanding of the etiology of schizophrenia.

Paper ID: 15

Robust Feature Extraction by Sample Pattern Embedding of single-cell RNA-seq data

Qianqian Shi[1], Chuanchao Zhang[2], Luonan Chen[3*]*

[1]College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

[2]Shanghai Institutes for Biological Sciences, China

[3]Osaka Sangyo University, China

Single-cell RNA-seq enables the quantitative characterization of cell types based on global transcriptome profiles, and holds enormous potential for both basic biology and clinical applications. However, scRNA-seq suffers from higher noise, unclear cell type and molecular regulatory. Hence, Molecular regulatory identification and characterization of cell types requires robust and accurate computational meth¬ods for Single-cell RNA-seq analysis. In the paper, we proposed a novel framework of Robust Feature Extraction (RFE), to identify the biological feature and characterize the Molecular regulatory heterogeneities in the cell levels, by embedding sample pattern. In particular, RFE have the robust performance for the single-cell RNA-seq data where there are the different missing values. To validate the utility of RFE, we analyze RFE on the six single-cell RNA-seq datasets, and found that RFE was capable to effectively capture the better features, which not only have the biological meaning but also reflect the intrinsic cell types, than the state-of-the-art integrative methods, such as SC3, t-SNE, Seurat and sparseDC.

Paper ID: 16

Predicting microRNA-disease associations based on microRNA structural and functional similarity network

Tao Ding, Jie Gao, Shanshan Zhu, Junhua Xu, Min Wu*

Jiangnan University, China

Increasing evidence indicates that microRNAs (miRNAs) are functionally related to the development and progression of various human diseases. Inferring more disease-related miRNAs can effectively promote disease biomarker detection for the treatment, diagnosis, and prevention of complex diseases. To improve the prediction accuracy of miRNA-disease association and capture more potential disease-related miRNAs, we construct a precise miRNA global similarity network (MSFSN) via calculating the miRNA similarity of secondary structures, families, and functions. We test new network on the classical algorithms: WBSMDA and RWRMDA through leave-one-out cross-validation. Eventually, AUCs of 0.8212 and 0.9657 are obtained, respectively. Also, MSFSN is applied to three cancers for breast neoplasms, hepatocellular carcinoma, and prostate neoplasms. Consequently, 82%, 76%, and 82% of the top 50 potential miRNAs for these diseases are respectively validated by miRNA-disease associations database miR2Disease and oncomiRDB. Therefore, MSFSN, a novel miRNA similarity network combining precise function network with global structure network of miRNAs, can be used to predict the associations between miRNAs and diseases in various models.

Paper ID: 18

Anti-CRISPRdb: a comprehensive online resource for anti-CRISPR proteins

Chuan Dong[1], Gefei Hao[2], Hongli Hua[1], Shuo Liu[1], Abraham Alemayehu Labena[1], Guoshi Chai[1], Jian Huang[1], Nini Rao[1], Fengbiao Guo[1]*

[1]University of Science and Technology of China, China

[2]Central China Normal University, China

CRISPR-Cas is a tool that is widely used for gene editing. However, unexpected off-target effects may occur as a result of long-term nuclease activity. Anti- CRISPR proteins, which are powerful molecules that inhibit the CRISPR–Cas system, may have the potential to promote better utilization of the CRISPR-Cas system in gene editing, especially for gene therapy. Additionally, more in-depth research on these proteins would help researchers to better understand the co-evolution of bacteria and phages. Therefore, it is necessary to collect and integrate data on various types of anti-CRISPRs. Herein, data on these proteins were manually gathered through data screening of the literatures. Then, the first online resource, anti-CRISPRdb, was constructed for effectively organizing these proteins. It contains the available protein sequences, DNA sequences, coding regions, source organisms, taxonomy, virulence, protein interactors and their corresponding three-dimensional structures. Users can access our database at http: //cefg.uestc.edu.cn/anti-CRISPRdb/ without registration. We believe that the anti-CRISPRdb can be used as a resource to facilitate research on anti-CRISPR proteins and in related fields.

Paper ID: 19

MicroRNAs in tsetse fly (Glossina morsitans) reveal novel insight to sleeping sickness

Zhiyuan Yang[1], Xi Zeng[2], Angel Tsz-Yau Wan[2]*

[1]Hangzhou Dianzi University, China

[2]The Chinese University of Hong Kong, Hong Kong

Tsetse fly (Glossina morsitans, GMR) is the main vector of disease trypanosomiasis causing up to 70 million deaths in Africa. MicroRNA (miRNA) is a kind of evolutionarily conserved small non-coding RNAs, playing critical roles in various biological process in the animal. To date, a global investigation of miRNA in tsetse fly and their contribution to trypanosomiasis is poorly understood. In this study, an EST-based homolog search is applied to identify potential miRNA in the tsetse fly. Target prediction and function annotation of identified miRNA were carried out to investigate the role of regulated miRNA in GMR. By using a series of comparative tools, we have identified six conserved and ten non-conserved miRNAs in the tsetse fly. The size, origin and predicted targets of identified miRNA were described. A total of 13 targeted 3'-UTRs were found in Glossina morsitans and the network was illustrated. In Conclusion, several miRNAs and complex population of small RNAs in Glossina morsitans suggest important roles for these non-coding RNAs in diverse sleeping sickness disease.

Paper ID: 20

Multi-scale modeling reveals angiogenesis-induced drug resistance in brain tumor and predicts a synergistic drug combination targeting EGFR and VEGFR pathways

*Weishan Liang, Xiaoqiang Sun**

Sun Yat-Sen University, China

Background

Experimental studies have demonstrated that both the extracellular vasculature, microenvironment and intracellualr molecular network (e.g. epidermal growth factor receptor (EGFR) signaling pathways) are essentially important for brain tumor growth. Some drugs have been developed to inhibit the EGFR signaling pathways. However, how does angiogenesis affect the response of tumor cells to the drug treatment has rarely been mechanistically studied. Therefore, a multiscale model is required to investigate such complex biological systems that contain interactions and feedbacks among multi-levels.

Results

In this study, we developed a single cell-based multi-scale spatio-temporal model to simulate more realistic vascular tumor growth and drug response, based on VEGFR signaling pathways, EGFR signaling pathway and cell cycle as well as several microenvironmental factors that determine cell fate switches in a temporal and spatial context.

The simulation reconstructed an evolving profile of vascular tumor growth, demonstrating the dynamic interplay between angiogenesis and various types of tumor cells (e.g., migrating, proliferating, apoptosis and quiescent cells). Incorporating EGFRI treatment effect, the model showed an interesting phenomenon that the survival rate of tumor cells decreased in the early stage but rebound in a later stage, revealing the emergence of drug resistance. Moreover, we revealed the critical role of angiogenesis in the acquired drug resistance, since inhibiting blood vessels' growth using VEGFR inhibitor prevented the recovery of survival rate of tumor cells in the later stage.

We further investigated the optimal timing of combing VEGFR inhibition with EGFR inhibition and predicted that the drug combination targeting both EGFR pathway and VEGFR pathway has a synergistic effect. The experimental data validated the prediction of drug synergy, confirming the effectiveness of our model.

Conclusions

The developed multiscale model revealed angiogenesis-induced drug resistance mechanisms of brain tumors to EGFRI treatment, and predicted a synergistic drug combination targeting both EGFR and VEGFR pathways with optimal combination timing. This study explored mechanistic and functional mechanisms of angiogenesis underlying tumor growth and drug resistance, which advances our understanding of novel mechanisms of drug resistance and provides implications for designing more effective cancer therapies.

Paper ID: 22

A deep learning framework for identifying essential proteins by integrating multiple sources of biological information

Min Zeng[1], Min Li[1], Zhihui Fei[1], Fangxiang Wu[2], Yaohang Li[3], Yi Pan[4], Jianxin Wang[1]*

[1]School of Information Science and Engineering, Central South University, China

[2]*University of Saskatchewan, Canada*
[3]*Old Dominion University, United States*
[4]*Georgia State University, United States*

Computational methods including centrality and machine learning-based methods have been proposed to accurately identifying essential proteins to understand the minimum requirements of the survival and evolution of a cell. In centrality methods, researchers are required to master prior knowledge to design a satisfied score function, yet a score function is usually not sufficiently expressive to capture the complexity of biological information. In machine learning-based methods, some selected biological features cannot represent the complete properties of biological information and there is lack of a computational framework to automatically select features. To tackle these problems, we propose a deep learning framework to automatically learn biological information features without prior knowledge. We use node2vec technique to automatically learn richer representation of PPI network topologies than a score function. Bidirectional long short term memory cells are applied to capture non-local relationships in gene expression data. For subcellular localization information, we exploit a high dimensional indicator vector to characterize their feature. To evaluate the performance of our method, we tested it on PPI network of S. cerevisiae. Our experimental results demonstrate that our method outperforms traditional centrality methods including DC, BC, CC, EC, NC, LAC, PeC and WDC. It also outperforms machine learning methods including SVM, decision tree, random forest and adaboost. To explore which of the three sources of biological information is the most vital element, we conduct an ablation study by removing each component. Our results show that the PPI network embedding contributes most to the improvement. In addition, gene expression data and subcellular localization information are also helpful to improve the performance in identifying essential proteins.

Paper ID: 23

Extracting Predictors for Lung Adenocarcinoma Based on Granger Causality Test and Stepwise Character Selection

*Xuemeng Fan, Xuqing Tang**
Jiangnan University, China

In order to extract several genes as predictors to classify tumors vs normal samples, a novel approach was proposed based on Granger Causality test and stepwise character selection, with the object of maximizing classification precision and minimizing number of predictors. The approach could be performed in three steps as follows in this paper. Firstly, the diff-genes were obtained by analyzing differential expression. Three datasets that the gene expression profiles, methylation data and miRNA expression data were all taken into account, owing to the dynamics of molecular expression, yet each single dataset just reflected the expression level at a static insight. Furthermore, the feature-genes were identified by network degree analysis on the basis of three gene-gene interactions which is the co-location interaction, the shared-protein domain interaction and physical interaction. Finally, Granger Causality test and Pearson correlation test based on the interaction network were utilized to remove “dependent-genes” from feature-gene set, and a stepwise character selection algorithm based on Random Forest classification model was further constructed. We added one gene into set

for each time and tried to delete old members in case they made no contribution to precision improvement. In our experiment, only 6 genes were exacted as resulting predictors, including TOP2A, GRK5, SIRT7, MCM7, EGFR, COL1A2. Robustness of this approach was validated by applying this 6-predictor-model into 6 independent datasets from GEO database. High precisions (ranged from 95.3% to 100%) indicated that our method was useful to classify patients and healthy individuals, which was also helpful in shortening the diagnosis time in clinical medicine.

Paper ID: 24

Identification subtypes of Non–Small Cell Lung Cancer

*Yuxuan Zhou, Fuyan Hu**

Wuhan University of Technology, China

As one of the most common malignancies in the world, lung cancer is difficult to be discovered or cured at the current medical level. The good news is that precision medicine provides new insights into the treatment of cancer. The strategy of precision medicine is that treatment is tailored to each patient's needs. To realize precision medicine for lung cancer, subtyping lung cancer plays an important role in revealing the pathological mechanism of the disease and performing specific treatment. This study started with human genome expression data and used statistical methods to classify lung cancer. First, we used the combination of K-means clustering and T statistical tests to reduce the dimension of the gene space, and then obtained genes that were differentially expressed in lung cancer and normal samples. Then, using these differentially expressed genes, the samples were mapped by the SOM neural network to obtain a link network between the samples. Next, four lung cancer subtypes were identified using spectral clustering methods. Afterward, significant differences in survival between the four subtypes were observed (p=0.0056). Finally, the results of function and pathway enrichment analysis for each subtype's representative genes showed that four subtypes had different pathological mechanisms. These four subtypes provide foundation for subtype-specific therapeutic of lung cancer.

Paper ID: 26

PTPD: Prediction of Therapeutic Peptides by Deep Learning

Chuanyan Wu, Rui Gao, Yusen Zhang*

Shandong university, China

In recent years, with the application of therapeutic peptides in the treatment of disease, many biologists spent a great deal of time and effort to verify various functional peptides from a large number of peptide sequences. In order to reduce workload and improve the efficiency of identification of functional proteins, we propose a deep-learning-based model to identify functional proteins (PTPD). At first, we divide the original peptide sequences into k-mers with a sliding window. Based on the co-occurrence information of k-mers, we obtain embedding vectors of all k-mers through Word2Vec. To get the feature map, three kinds of filters are used in the convolutional layer together with dropout layers and max-pooling layers. To learn complex high-level relationships, the results are concatenated together followed by fully-connected dense layers of

rectified linear units (ReLU) with dropout to avoid over-fitting. Finally, sigmoid function is used to generate the final classification probabilities. To verify the performance of PTPD, we performed the PTPD on four ACPs datasets and two virulent proteins datasets, which got accuracies of 99.52%, 93.08%, 97.78%, 91.11%, 100%, and 91.18%, respectively. It is anticipated that PTPD can be effective to identify the novel therapeutic peptides and maybe helpful to design the therapeutic peptides.

Paper ID: 28

Predicting stage-specific cancer related genes and their dynamic modules by integrating multiple datasets.

Aouiche Chaima, Bolin Chen, Xuequn Shang*
Northwestern Polytechnical University, China

The majority of complex diseases progression have not been detected accurately by the mechanism of staging evolution.
Exploring the dynamics of these stages through biological modules from integrated omics data has gained a new insight into genomic and clinical research.
Although many intensive studies have been developed on this topic, less is known about the specific cancer related genes that are associated with individual stages.
Additionally, critical direct or indirect relations between those identified cancer related genes across multiple datasets.
Thus, we have proposed powerful and versatile approaches to identify stage-specific cancer genes and discover their related dynamic modules, based on two genomic datasets and a human PPI network.
The discovered modules and their specific-signature genes were significantly enriched in some relevant known pathways.
As a further step, we show the particular dynamic evolution of these clinical-stage by revealing relationships between their enriched pathways based on the overlap between their annotated genes.
As a result, a significant pathway network has been built, which not only help us to understand the functional evolution of complex diseases, but also the clinical management by selecting the optimum treatment regimens for patients and the appropriate drugs.

Paper ID: 30

MD-SVM: A novel SVM-based algorithm for the motif discovery of transcription factor binding sites

Jialu Hu, Jingru Wang, Xuequn Shang*
Northwestern Polytechnical University, China

Transcription factors (TFs) play important roles in the regulation of gene ex- pression. They can activate or block transcription of downstream genes in a manner of binding to specific genomic sequences. Therefore, motif discovery of these binding preference patterns is of central significance in the understanding of molecular regulation mechanism. There are many algorithms for the identifi-

cation of transcription factor binding sites. However, it remains a challengeable problem. Here, we proposed a novel motif discovery algorithm based on support vector machine (MD-SVM) to learn a discriminative model for TF binding sites. MD-SVM firstly obtains position weight matrix (PWM) from a set of training datasets. Then it translates the MD problem into a computational framework of multiple instance learning (MIL). It was applied to several real biological datasets. Results show that our algorithm outperforms MI-SVM in terms of both accuracy and specificity.

Paper ID: 31

PWCDA: A New and Efficient Method for Predicting circRNA-disease Associations

Xiujuan Lei[1], Zengqiang Fang[1], Luonan Chen[2], Fangxiang Wu[3]*

[1]Shaanxi Normal University, China

[2]Chinese Academy of Sciences, China

[3]University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

CircRNAs have particular biological structure and have proven to play important roles in diseases. It is time-consuming and costly to identify circRNA-disease associations by biological experiments. Therefore, it is appealing to develop computational methods for predicting circRNA-disease associations. In this study, we propose a new computational model (called PWCDA) to predict circRNA-disease. Firstly, we picked out 592 circRNA-disease associations from circR2Disease. Based on these 592 associations, we matched 83 diseases and 541 circRNAs respectively. Secondly, according to disease-related gene annotations and circRNA related gene ontology, we obtain disease functional similarity network and circRNA functional similarity network, respectively. Considering that there are lots of similarity scores missing, we calculate the Gaussian Interaction Profile GIP kernel similarity for diseases as well as circRNAs. Then, we integrate disease/circRNA functional similarity scores and their related GIP kernel similarity scores. Then we build a heterogeneous network of three sub-networks. Disease functional similarity network, circRNA functional similarity network and circRNA-disease association network. Finally, we calculate an association score for each circRNA-disease pair based on the paths connecting them in the heterogeneous network and use the threshold method to determine whether a pair of circRNA-disease is associated. We adopt leave one out cross validation (LOOCV) and five-fold cross validations to investigate the performance of our proposed method. In addition, three common diseases, Breast Cancer, Gastric Cancer and Colorectal Cancer, are used for case studies. Case studies illustrate the reliability of our computational model in terms of different validation measures, which indicates PWCDA is a useful and reliable bioinformatics tool to infer potential circRNA-disease associations.

Paper ID: 32

Detecting the stable point of therapeutic effect of chronic myeloid leukemia based on dynamic network biomarkers

Junhua Xu[1], Min Wu[1], Shanshan Zhu[1], Jinzhi Lei[2], Jie Gao[1]*

[1]Jiangnan University, China

[2]Zhou Pei-Yuan Center for Applied Mathematics, China

For chronic myeloid leukemia (CML), most researches are currently focused on the treatment of disease. There is a few researches on the development of patients after drug treatment. In this paper, we show a sample-specific strategy based on individual-specific dynamic network biomarkers (DNB). Using the CML microarray data and based on DNB criteria, the DNB including 250 genes are chosen and therapeutic effect index (TEI) is constructed for the detection of individual samples. Then, it is obtained that the best stable point of TEI is 1 month. Through functional analysis for DNB, some genes are confirmed as key genes to affect the growth of the CML patients' condition.

Paper ID: 33

Diagnosing pre-disease state of influenza A disease based on dynamical network biomarkers

Shanshan Zhu, Jie Gao, Junhua Xu, Min Wu, Tao Ding*
Jiangnan University, China

Evidences have shown that there is state denoted as pre-disease state in which the disease progression shows strongly fluctuations. The living organism performs a smooth change with time and conditions while in other states. And after this critical transition point, the disease progression drives the state from normal to disease rapidly. Considering differences in expression with time, this paper provides a method that filters optimal dynamical network biomarkers (DNB) by seek out the clusters K at different time points. Then, an indicator with dynamical network biomarkers (DNB) is defined as early warning index (EWI) for detecting pre-disease state. Combined with microarray data of influenza A disease, 95 genes are selected as DNB in which 22 genes are confirmed to be related to influenza A, and the state from 4th(29 hours) to 5th(36 hours) time period is denoted as pre-disease state. Taken together, these findings show that the discovered DNB is relevant with experience data, which can illustrate the effectiveness of our method.

Paper ID: 34

Prioritizing type 2 diabetes genes by weighted PageRank on bilayer heterogeneous networks

*Haixia Shang, Zhiping Liu**
Shandong University, China

The prevalence of diabetes mellitus has been increasing rapidly in recent years. Type 2 diabetes makes up about 90% cases of diabetes. The interacting mixed effects of genetics and environments build interpretable possible pathogenesis. Thus, finding the causal disease genes is crucial in clinical diagnosis and medical treatment. Currently, network-based method becomes a powerful tool of systematically analyzing complex diseases, such as the identification of candidate disease genes. In this paper, we propose a bioinformatics framework of prioritizing type 2 diabetes genes by leveraging the modified PageRank algorithm on bilayer biomolecular networks consisting an ensemble gene-gene regulatory network and an integrative protein-protein interaction network. We specifically weigh the networks by differential mutual information for measuring the context specificities between genes and between proteins by transcriptomic and proteomic datasets. After formulating the network into two components of known disease genes and the other normal healthy genes, we rank the diabetes genes and other genes by bringing the order in the bilayer network via

improved PageRank. We conclude that these known disease genes achieve significantly higher ranks compared to the normal genes, and the ranks are robust and consistent in multiple validation scenarios. These high-ranked genes are identified to perform relevant dysfunctions of type 2 diabetes risks.

Paper ID: 35

CNAPE: A Software for Copy Number Alteration Prediction from Gene Expression in Human Cancers

*Quanhua Mu, Jiguang Wang**

The Hong Kong University of Science and Technology, Hong Kong

Copy number alteration (CNA), the abnormal number of copies of genes, chromosomal-arms or even whole chromosomes, is commonly observed in human cancers. And CNA is believed to play an essential role in cancer progression and disease diagnosis. Current high-throughput CNA detection methods include array comparative genomic hybridization (aCGH) and genomic sequencing. However, both thechnologies are relatively expensive and require DNA samples at microgram level, which is not achievable in certain occasions such as tissue biopsies or single-cell genomics. Here we proposed an alternative computational method—CNAPE to infer somatic CNAs using RNA sequencing data. The machine learning model was trained on the trancriptomic profile with matched copy number data of 9,740 cancers from The Cancer Genome Atlas, and then applied to an independent bulk RNA sequencing dataset. CNAPE achieved >90% accuracy in the prediction of arm-level CNAs. Specificlly, for chromosome 1p/19q co-deletion, an essential molecular biomarker for brain tumor diagnosis, our model achieved an AUC of 0.992, superior to the previously published method. For focal CNA, the accuracy varied from 60% to 90%, depending on the sizes and locations of the genes. In addition, CNAPE was developend as an open-access software for public usage.

Paper ID: 36

ODAE: Ontology-based systematic representation and analysis of drug adverse events and its usage in study of adverse events given different patient age and disease conditions

Hong Yu[1,2], Solomiya Nysak[3], Noemi Garg[3], Edison Ong[4], Yongqun He[5*]*

[1]Guizhou Provincial People's Hospital, China

[2]Guizhou University Medical College, China

[3]College of Literature, Science, and the Arts, University of Michigan, United States

[4]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, United States

[5]University of Michigan, United States

Background

Drug adverse events, or called adverse drug events (ADEs), are ranked one of the leading causes of mortality. The Ontology of Adverse Events (OAE) has been widely used for adverse event AE representation, standardization, and analysis. OAE-based ADE-specific ontologies, including

ODNAE for drug-associated neuropathy-inducing AEs and OCVDAE for cardiovascular drug AEs, have also been developed and used. However, these ADE-specific ontologies do not consider the effects of other factors (e.g., age and drug-treated disease) on the outcomes of ADEs. With more ontological studies of ADEs, it is also critical to develop a general purpose ontology for representing ADEs for various types of drugs.

Results

Our survey of FDA drug package insert documents and other resources for 224 neuropathy-inducing drugs discovered that many drugs (e.g., sirolimus and linezolid) cause different AEs given patients' age or the diseases treated by the drugs. To logically represent the complex relations among drug, drug ingredient and mechanism of action, AE, age, disease, and other related factors, an ontology design pattern was developed and applied to generate a community-driven open-source Ontology of Drug Adverse Events (ODAE). The ODAE development follows the OBO Foundry ontology development principles (e.g., openness and collaboration). Extending the OAE and NDF-RT ontology, ODAE has represented various AEs associated with the over 200 neuropathy-inducing drugs. Examples are provided to illustrate how ODAE represents various ADE-related complex entities and their relations given different conditions. ODAE is now deposited in the Ontobee for browsing and queries. SPARQL queries can also be developed to query the ODAE knowledge base for different questions. For example, a simple SPARQL query of ODAE identified 35 neuropathy-inducing drugs that may be used to treat respiratory diseases.

Conclusions

The ODAE ontology provides a general design pattern for ADE representation and represents a large number of ADEs given different conditions. ODAE can serve as a knowledge base of ADEs and be used for data mining to address different scientific questions. Furthermore, the novel ODAE can be used as a robust platform for semantic and logic representation and study of ADEs of more drugs in the future.

Paper ID: 37

A voting mechanism-based linear epitope prediction system for the host-specific Iridoviridae family

Tunwen Pai

Department of Computer Science and Engineering, National Taiwan Ocean University, Taiwan

The Iridoviridae family is divided into two subfamilies: Alphairidovirinae includes Lymphocystivirus, Ranavirus (GIV), and Megalocystivirus (TGIV), which infect vertebrate hosts and Betairidovirinae includes Iridovirus and Chloriridovirus, which infect invertebrate hosts. Grouped Iridoviridae subfamilies possess host-specific characteristics, which can be considered as exclusive features for identifying effective linear epitopes (LEs). A voting mechanism-based LE prediction system was proposed to analyze any two differently clustered pathogen groups, allowing both conserved and exclusive LEs to be identified simultaneously. In this study, the advantages of undocumented cross-infection between vertebrate and invertebrate host species of the Iridoviridae family were applied to reevaluate the impact of LE prediction. Using five renowned LE prediction systems and exclusive features, endorsed LE candidates with a minimum length requirement could be identified for each subfamily by various prediction systems. Furthermore, surface structural

characteristics of identified conserved and exclusive LE candidates were confirmed through structural alignment analysis. To validate the predicted LEs, ELISA assays were performed to identify vertebrate host-specific LEs for both GIV and TGIV, and the experimental results showed that predicted LEs were reflected in high antigenicity responses for specific grouper species. Therefore, it demonstrates that the proposed system provides an effective approach for in silico LE prediction prior to vaccine development, and it is especially powerful for analyzing antigen sequences with exclusive features between two clustered groups.

Paper ID: 38

The Cis-trans Binding Strength Defined by Motif Frequencies Facilitates Statistical Inference of Transcriptional Regulation

*Von Weber , Lei Li**

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

Gene functions through transcription. Transcription regulation allows the organism to respond to a variety of intra- and extra-cellular signals, maintain biological differentiation and evolutionary development. Many biotechnology and methodology in computational biology were developed to this end. Lin Wang, Chao Cheng proposed BASE (binding association with sorted expression) method, to infer the activity changed transcript factors (TFs) with gene expression data and experimental TF binding affinity data. In order to reduce the dependence with experiment and boost the availability, we define a new measurement of TF binding affinity based on motif analysis. The modified BASE method just uses the gene expression data as input and leaves other work done in silico. In the study to explore the effects of breast cancer treatment and toxicity of nanoparticle 〖[Gd@C_82 〖(OH)〗_22]〗_n, with the modified BASE method, we discover many meaning TFs and conclude that the nanoparticle treats breast cancer by trying to reestablish the crucial hormone receptors of breast tumor to restore cells' physiological functions and by strengthening the tumor suppressors. Furthermore, in the study to explore the effects of polyunsaturated fatty acids EPA&DHA diets to mouse small intestine cells, we find EPA&DHA diets promote the health of mouse in many aspects, such as reducing cholesterol biosynthesis, reducing inflammation, reducing angiogenesis, enhancing insulin sensitivity and promoting adipogenesis.

Paper ID: 39

Biosystems study of the molecular networks underlying Alzheimer's disease: differences between wild-type and Atp11b-knockout mice

*Jiao Wang, Fangfang Zhou, Jie Li, Qian Li, Weihao Li, Fangfang Ma, Tieqiao Wen**

Laboratory of Molecular Neural Biology, School of Life Sciences, Shanghai University, China

Alzheimer's disease (AD) is a neurodegenerative disease with complex pathogenesis, which is characterized by learning and memory deficits and usually involves multiple systems. In our previous study, we showed that ATP11B deficiency induces learning and memory impairments, which are the functional neurological disorders of AD patients. Moreover, we found that ATP11B expression is decreased in a transgenic mouse model of AD; however, the mechanism underlying

ATP11B-induced pathogenesis and development of AD has not yet been established. To investigate the involvement of ATP11B in AD, differentially expressed genes (DEGs) between wild-type (WT) and Atp11b-knockout (KO) mice were identified, and the overlapping DEGs and AD-related genes were analyzed. The results reveal that ATP11B plays a pivotal role in AD via the immune response and metabolic regulation. Moreover, the majority of overlapping genes were enriched in the regulation of gamma-aminobutyric acid (GABA) secretion and synaptic transmission, which contribute to the process of AD development. Furthermore, molecular interaction analysis of the DEGs indicates that the hub gene is S100a9, which plays crucial roles during AD progression, suggesting that ATP11B may be involved in the development of AD. Subsequently, biological experiments were performed to further confirm the accuracy of the molecular interaction network. These results reveal that ATP11B may play an important role in AD pathogenesis. The present study may shed light on a possible mechanism of ATP11B in AD, thereby providing a theoretical basis for novel AD therapies.

Paper ID: 40

RISE: a database of RNA interactome from sequencing experiments

Qiangfeng Zhang
Tsinghua University, China

We present RISE (http://rise.zhanglab.net), a database of RNA Interactome from Sequencing Experiments. RNA-RNA interactions (RRIs) are essential for RNA regulation and function. RISE provides a comprehensive collection of RRIs that mainly come from recent transcriptome-wide sequencing-based experiments like PARIS, SPLASH, LIGR-seq, and MARIO, as well as targeted studies like RIA-seq, RAP-RNA and CLASH. It also includes interactions aggregated from other primary databases and publications. The RISE database currently contains 328,811 RNA-RNA interactions mainly in human, mouse and yeast. While most existing RNA databases mainly contain interactions of miRNA targeting, notably, more than half of the RRIs in RISE are among mRNA and long non-coding RNAs. We compared different RRI datasets in RISE and found limited overlaps in interactions resolved by different techniques and in different cell lines. It may suggest technology preference and also dynamic natures of RRIs. We also analyzed the basic features of the human and mouse RRI networks and found that they tend to be scale-free, small-world, hierarchical and modular. The analysis may nominate important RNAs or RRIs for further investigation. Finally, RISE provides a Circos plot and several table views for integrative visualization, with extensive molecular and functional annotations to facilitate exploration of biological functions for any RRI of interest.

Paper ID: 41

Generalized Gene Co-Expression Analysis via Subspace Clustering Using Low-Rank Representation

Tongxin Wang[1], Jie Zhang[2] and Kun Huang[2]*
[1]Indiana University Bloomington, USA
[2]Indiana University School of Medicine, USA

Motivation:Gene co-expression network analysis (GCNA) helps to identify gene modules with potential biological functions and has become a popular method in bioinformatics and biomedical research. Most current GCNA algorithms use correlation to build gene co-expression network and identify gene modules with highly correlated genes. There is a need to look beyond correlation and identify gene modules using other similarity measures in order to find new biologically meaningful gene modules.
Results: We propose a new GCNA algorithm that could identify biologically meaningful gene co-expression modules with genes that are not all highly correlated. We use low-rank representation (LRR) to generate gene co-expression network and local maximal quasi-clique merger (lmQCM) to identify gene co-expression modules. We show that our method could identify gene modules with different biological functions and than current GCNA algorithms in different microarray datasets, as well as finding gene modules with prognostic values.

Paper ID: 42

Cell Lysate Microarray for Mapping the Network of Genetic Regulators for Histone Marks

Li Cheng[1], Jun-Biao Dai[2] and Sheng-Ce Tao[1*]*
[1]Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, China
[2]Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, CAS, China

Proteins, as the major executer for cell progresses and functions, its abundance and the level of post-translational modifications, are tightly monitored by regulators. Genetic perturbation could help us to understand the relationships between genes and protein functions. Herein, to explore the impact of the genome-wide interruption on certain protein, we developed a cell lysate microarray on kilo-conditions (CLICK) with 4,837 knockout (YKO) and 322 temperature-sensitive (ts) mutant strains of yeast (Saccharomyces cerevisiae). Taking histone marks as examples, a general workflow was established for the global identification of upstream regulators. Through a single CLICK array test, we obtained a series of regulators for H3K4me3, which covers most of the known regulators in S.accharomyces. We also noted that several group of proteins that are involved in negatively regulation of H3K4me3. Further, we discovered that Cab4p and Cab5p, two key enzymes of CoA biosynthesis, play central roles in histone acylation. Because of its general applicability, CLICK array could be easily adopted to rapid and global identification of upstream protein/enzyme(s) that regulate/modify the level of a protein or the posttranslational modification of a non-histone protein.

Paper ID: 43

Discovery of the substrate of deubquitinase USP9X by quantitative proteomics

Hu Zhou, Xiangling Chen, Chengli Yu and Jing Gao**
Shanghai Institute of Materia Medica, CAS, China

The X-linked deubiquitinase, USP9X, is implicated in multiple cancers by targeting various substrates. Increased expression of USP9X is observed in non-small cell lung cancer (NSCLC) and is correlated with poor prognosis. However, the molecular mechanism for USP9X regulating tumor

cell survival and tumorigenesis in NSCLC is less defined. In this study, TMT-based chemical labeling quantitative proteomic screening was applied to analyze A549 cells with or without USP9X RNA interference,resulting in a total of 7471 proteins identified. Bioinformatic analysis of the proteomic data suggested that Dual specificity protein kinase TTK is a potential substrate of USP9X. Further experimental evidences confirmed that USP9X stabilized TTK via direct interaction and deubiquitination of TTK. Moreover, knockdown of USP9X or TTK inhibited cell proliferation, migration and tumorigenesis, and the immunohistochemical analysis of clinical NSCLC samples showed that the protein expression levels of USP9X and TTK were significantly elevated and positively correlated in tumor tissues. In summary, our data demonstrated that the USP9X-TTK axis may play a critical role in NSCLC, and could be considered as the potential therapeutic target.

Paper ID: 44

Parameter sensitivity analysis for a stochastic model of mitochondrial apoptosis pathway

Xianli Chen[1], Xiaoguang Li[2], Wei Zhao[3], Qi Ouyang[4] and Tiejun Li[5]*

[1] Peking University, China

[2]College of Mathematics and Compute Science, Hunan Normal University, Changsha, China

[3]Center for Quantitative Biology and Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

[4]The State Key Laboratory for Artificial Microstructures and Mesoscopic Physics, Department of Physics, China

[5]LMAM and School of Mathematical Sciences, Peking University, Beijing, China

Understanding how gene alterations induce oncogenesis plays an important role in cancer research and may be instructive for cancer prevention and treatment. We conducted a parameter sensitivity analysis to the mitochondrial apoptosis model. Both a nonlinear bifurcation analysis of the deterministic dynamics and energy barrier analysis of the corresponding stochastic models were performed. We found that the parameter sensitivity ranking according to the change of the bifurcation-point locations in deterministic models and the change of the barrier heights from a living to death state of the cell in stochastic models are highly correlated. For the model we considered, in combination with previous knowledge that the parameters significantly affecting the system's bifurcation point are strongly associated with frequently mutated oncogenic genes, we conclude that the energy barrier height can be used as indicator of oncogenesis as well as bifurcation point. We provide a possible mechanism that may help elucidate the logic of cancer initiation from the view of stochastic dynamics and energy landscape. And we show the equivalence of energy barrier height and bifurcation-point location in determining the parameter sensitivity spectrum for the first time.

Paper ID: 45

Cancer development as a dynamical system

Jinzhi Lei

Tsinghua University, China

Cancer development is a complex process. However, if we omit the details of molecular level changes, we can consider the process of cancer development as a dynamical systems of abnormal cell growth. In the paper, we will introduce a general mathematical framework that model caner development through a dynamical system with heterogeneity in cell regeneration. Based on the model framework, we establish a computation model that reproduces the process of inflammation-induced tumorigenesis.

Paper ID: 46

The free energy cost of synchronization in caynobacterial post-transcriptional circadian clock

Dongliang Zhang[1], Yuansheng Cao[2], Yuhai Tu[3] and Qi Ouyang[1]*
[1] Peking University, China
[2]University of California San Diego, USA
[3]IBM W.J Research Center, USA

Synchronization is an important kind of function in biological rhythm and timekeeping. However, it's thermodynamic cost is still not clear. It's little known whether synchronization cost energy, or how the free energy is used. Here we provide models describing how energy is cost in caynobacterial post-transcriptional circadian clock, showing that it requires additional energy for a synchronized oscillation, no matter what the details are. Our results can help us understand the thermodynamic cost of general synchronization better.

Paper ID: 47

Concept and application of differential network model in single-sample analysis of biological big data

Tao Zheng
Key Laboratory of Systems Biology, CAS, China

The differential network model is a useful network-based approach to characterize the system change, e.g. the genotype change of biological system. This model characterizes the genes with changed expression as nodes, and the gene-pairs with changed co-expression as edges. Then any network centrality can be applied to indicate the association between the gene network change and phenotype change.

Due to the biological specificity, the single-sample analysis rather than conventional population-sample analysis is getting a lot of attraction in the era of biological big data. It is expected to capture the sample-specificity of biology with the support from the high-throughput omics technologies.

Thus, how to apply differential network model in single-sample analysis is widely required for improving both the interpretability and discrimination of genotype-phenotype association study, e.g. the precision medicine or personalized medicine. Here, this highlight firstly illustrates the concept of differential network model in single-sample analysis from the viewpoint of differential expression mean, differential expression variation, and differential expression co-variance. Then it introduces the wide applications of single-sample differential network model in predicting disease states, identifying cancer drivers and even recognizing rice-production associated miRNAs.

Paper ID: 48

Network Analysis of Single-cell RNA Sequencing Data based on cell-specific network

*Hao Dai, Lin Li, Tao Zeng and Luonan Chen**

Shanghai Institute of Biochemistry and Cell Biology, CAS, China

We present a new method that constructs a cell-specific network (CSN) for each cell from single-cell RNA-seq data. Based on CSN, we construct a network degree matrix to perform clustering, dimension-reduction and pseudo trajectory analyses on several published data sets. Compared with the traditional gene expression matrix, our method shows better performance in all analysis methods, which indicates the validity and meaningfulness of CSN.

Paper ID: 49

Modeling Endoplasmic Reticulum Network Maintenance in a Plant Cell

Congping Lin[1], Rhiannon White[2], Imogen Sparkes[3] and Peter Ashwin[2]*

[1] Huazhong University of Science and Technology, China

[2]University of Exeter, UK

[3]University of Bristol, UK

The endoplasmic reticulum (ER) in plant cells forms a highly dynamic network of complex geometry. ER network morphology and dynamics are influenced by a number of biophysical processes, including filament/tubule tension, viscous forces, Brownian diffusion, and interactions with many other organelles and cytoskeletal elements. Previous studies have indicated that ER networks can be thought of as constrained minimal-length networks acted on by a variety of forces that perturb and/or remodel the network. Here, we study two specific biophysical processes involved in remodeling. One is the dynamic relaxation process involving a combination of tubule tension and viscous forces. The other is the rapid creation of cross-connection tubules by direct or indirect interactions with cytoskeletal elements. These processes are able to remodel the ER network: the first reduces network length and complexity whereas the second increases both. Using live cell imaging of ER network dynamics in tobacco leaf epidermal cells, we examine these processes on ER network dynamics. Away from regions of cytoplasmic streaming, we suggest that the dynamic network structure is a balance between the two processes, and we build an integrative model of the two processes for network remodeling. This computational model produces quantitatively similar ER networks to those observed in experiments. We use the model to explore the effect of parameter variation on statistical properties of the ER network.

Paper ID: 50

Sequential data analysis based on constrained optimization by bivariate PCA

*Si Zhang, Tao Zeng and Luonan Chen**

Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, China

Evolution is an important branch of the biological sciences. However, these evolutionary data are often complex. For example, in the study of high-altitude adaptation of multi-tissue trans-altitude plateau animals, the data mainly includes three important variables: altitude, tissue, and individual. Due to data complexity, principal component analysis (PCA) cannot well distinguish the differences between each organization among altitudes. Therefore, we use bivariate principal component analysis. This model indicates the relationship between variables 2 and 3 by reducing the impact of variable 1, and imposing constraints on variables 2 and 3. In simulation experiment, our model tend to be high correlated with true pattern. Using the model, we can discover each tissue's specific pattern, which facilitates us to better understand the role that different organizations may play in the process of phenotype adaptation.

Paper ID: 51

Predicting Conserved Regions in Protein Sequences Using Equivalence Classes

Jingsong Zhang, Tao Zeng and Luonan Chen*

Institute of Biochemistry and Cell Biology, CAS, China

The identification of conserved regions in biological sequences is essential to predict structural and functional regions such as motifs, domains and genes. The ab initio calculation predicts conserved regions from the amino acid sequences that contain new conserved regions, which have not been characterized or studied yet. However, existing ab initio methods suffer from low prediction accuracy. Moreover, they often fail in prediction when feeding long protein sequences. Inspired by the successful application of the equivalence relationship of patterns in sequential pattern mining, we propose a novel ab initio approach called ProRegion (Protein Region predictor) to predict protein conserved regions. ProRegion utilizes the intrinsically conserved features of equivalence classes and explores the equivalence relationship of patterns within equivalence classes to discover the equivalence patterns, which are compact representation of equivalence classes. ProRegion subsequently explores the equivalence patterns as the cores of protein conserved regions. Our experiments demonstrate that ProRegion effectively predicts conserved regions from protein sequences with a wide range of lengths, especially from long protein sequences.

Paper ID: 52

Network clustering of Single-cell RNA Sequencing Data

*Lin Li and Luonan Chen**

CAS, China

Single-cell RNA-seq enables the quantitative characterization of cell types based on global transcriptome profiles. We present a new method that constructs a network for each cell from single-cell RNA-seq data. Our method defines an edge as an irreducible statistical dependency between gene expression profiles. We suggest that the presence of such statistical dependencies is likely to identify direct regulatory interactions. And then we get the invariant measure of the network of each cell to perform clustering, dimension-reduction on several published data sets. And we can also get network entropy based the network of each cell to accurate estimation of differentiation

potency .Compared with the traditional gene expression matrix, our method shows better performance in all analysis method.

Paper ID: 53

Analysis of Hyperdysregulatory network in Human with Pancancer Data

Pingyang Wang[1], Lina Lu[1*], Tao Zeng[2*] and Luonan Chen[2*]*

[1]Shanghai Institute of Biochemistry and Cell Biology, China

[2]Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, CAS, China

Tumorigenesis generally results not from individual molecules but from dysfunction of the relevant system or network, which dynamically changes with time and conditions. We present hyperdysregulation theory drawing on the previous study of hypermutation through expression data analysis of approximately 7000 tumor samples obtained from the TCGA database, including 16 tumor types. We construct single sample networks (SSN) based on molecular expressions using our lab's published method. The SSN for each sample is constructed based on statistical perturbation analysis of this sample against a group of given control samples. In our definition, a sample is considered as a hyperdysregulated sample if its amount of edges in SSN exceeds preseted threshold. Thus, we count amount of edges in every SSN and find out dysregulation samples. We find that the edge counts in samples diagnosed as Kidney Chromophobe (KICH) are significantly higher than those in other tumor types. In general, each type of tumors shows significant difference in their distribution. Following that, we try to analyze the correlation between the edges counts and the overall survival which can be used as a network biomarker of disease prognosis in the future. Meanwhile, we use node's information in SSN to divide hyperdysregulated samples into different subtypes in each tumor types. Then we search for differences and connections between these subtypes and do analysis combined with individual clinical information. Through the establishment of this theory, we will get a macroscopic metric in individual-specific expression network to measure individual cancer condition.

Paper ID: 54

Analysis of molecular mechanisms for EGFR-TKI resistance in NSCLC based on transcriptome data

Tang Shijie and Chen Luonan[*]

Shanghai Institute of Biochemistry and Cell Biology, China

Non-small cell lung cancer (NSCLC) is the major subtype of lung cancer and accounts for about 85% of all lung cancers. With deeper digging into the oncogenesis and progression mechanisms of NSCLC, epidermal growth factor receptor (EGFR) become one of the landmark targets of NSCLC therapy, however, drug resistance occurs in almost all patients with initial dramatic responses to EGFR-TKI within 6-12 months. This study aims to discover the EGFR-TKI resistance molecular mechanisms in NSCLC with time series RNA-Seq data from TKI treatment PC9 cells. 27 samples were analyzed in this study including control PC9 cells, PC9 cells with TKI treatment for 2 hours, 6 hours, 12 hours, 24 hours, 48 hours, P4, P8 and P16, each time point with 3 samples. Dynamical

network biomarkers (DNB) analysis was performed on our time series RNA-Seq data in order to detect the critical time point during the drug treatment process. TKI treatment for 6 hours and P4 were detected to be two possible critical time points related with drug resistance in the treatment process. Then, pathway enrichment analysis was applied on differentially expressed genes between time points before and after the detected critical time point and ISMARA was performed to identify key transcription factors driving expression changes. We found that genes and pathways related to cell cycle were down-regulated after TKI treatment for 6 hours and up-regulated after P4 time point. What's more, transcription factors related to cell cycle were inactived after TKI treatment for 6 hours but activated after TKI treatment for 6 hours as well. These results indicate that time point of TKI treatment 6 hours may be the drug efficacy time point and P4 may be the drug resistance time point.

Paper ID: 55

Photocatalytic protein damage of silver nanoparticles circumvents bacterial stress-response and thus abolish multidrug resistance

Tianyuan Shi, Qiuxia Wei, Gong Zhang, Xuesong Sun* and Qing-Yu He**

Jinan University, China

Silver nanoparticles (AgNPs) are known for its broad-spectrum antibacterial properties, especially against those antibiotic-resistant bacteria. However, the real bactericidal mechanism of AgNPs remains unclear. In this study, we found that the bactericidal ability of AgNPs are induced by light. Unlike the previous postulations, visible light is unable to release silver ions released from AgNPs, nor to induce ROS in Escherichia coli. In contrast, we found that the light excited AgNPs induced protein aggregation in a concentration dependent manner in E. coli, indicating that the bactericidal ability of AgNPs lays on the light-catalyzed oxidation of cellular proteins via direct binding to proteins. This was verified by fluorescence spectra. iTRAQ based proteomics revealed that proteins in E. coli was oxidative damaged by the light-excited AgNPs. In sum, AgNPs absorbs the energy of light and transfer it to the proteins, leading to the oxidation of proteins and thus promotes the death of the bacteria. This direct and physical mechanism is unlikely to be avoided by any known drug-resistance mechanisms of the cells and thus can serve as a last resort against the drug-resistance. This mechanism also provided a practical hint on the antimicrobial application of the AgNPs – exposing it to the light.

Paper ID: 56

Profiling and functional analysis of circular RNAs in acute promyelocytic leukemia and their dynamic regulation during all-trans retinoic acid treatment

*Shufen Li, Yunlin Ma, Yun Tan, Xuefei Ma, Ming Zhao, Bing Chen, Rongsheng Zhang, Zhu Chen and Kankan Wang**

State Key Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, China

Background: Circular RNAs (circRNAs) are a novel class of powerful regulators in gene expression

and participate in the pathogenesis of many diseases, including cancer. However, little is known about the roles of circRNAs in the development and treatment of acute promyelocytic leukemia (APL).

Methods: We performed two independent ribosomal RNA-minus RNA-sequencing (Ribo-minus RNA-seq) experiments, one with and one without RNase R treatment, to identify circRNAs in APL cells before and 24 and 48 hours after ATRA treatment in APL patient-derived NB4 cells. We also retrieved public available Ribo-minus RNA-seq datasets from three leukocyte types: hematopoietic stem cells (CD34+), naive B cells (CD19+) and neutrophils. STAR and DCC were used to detect circRNAs, and TopHat2 and Cufflinks were employed to calculate the FPKM value of each host gene. We characterized and verified these circRNAs in NB4 and one candidate circRNA was identified from the significantly deregulated circRNAs to perform functional and mechanistic experiments.

Results: We identified a total of 4,313 circRNAs, including 1,098 newly identified circRNAs. Detailed analysis showed that circRNAs expressed in APL cells were mostly exon-derived, not by-products during splicing and could be distinguished from hematopoietic stem cells, neutrophils and lymphocytes. The true presence and stability of circRNAs were verified both in NB4 cells and primary APL patient samples. Moreover, we conducted a time-series analysis of circRNAs on ATRA-treated NB4 cells and uncovered 508 circRNAs with dynamic expression during ATRA treatment, including 246 up-regulated and 262 down-regulated. Further evidence demonstrated that the majority of circRNAs were regulated independently of their host linear mRNAs. Detailed functional experiments demonstrated that circ-HIPK2, one of the differentially expressed circRNAs, significantly influenced ATRA-induced differentiation of APL cells. Further mechanistic studies revealed that circ-HIPK2 was located in cytoplasm and served as a sponge for differentiation-associated miR-124-3p. Finally, circ-HIPK2 expression in APL patients was significantly lower than that in normal peripheral mononuclear cells and other subtypes of AML, indicating its potential role as an APL biomarker.

Conclusion: Our study identified a large number of dynamically regulated circRNAs during ATRA-induced APL cell differentiation, thus providing an important basis for further studies addressing their function and suitability as biomarkers. Moreover, we determined the biological function and mechanisms of circ-HIPK2 in the regulation of APL differentiation. Further investigation is warranted to uncover more circRNAs that participate in APL occurrence and development.

Paper ID: 57

scLRTD : A novel low rank tensor decomposition method for single-cell RNA-sequencing data imputing

Zhijie Ni, Dingjie Wang, Xiao Zheng and Xiufen Zou[*]

Wuhan University, China

With the successful application of single-cell sequencing technology, a large number of single-cell RNA-sequencing (scRNA-seq) data have been generated, which enables researchers to study heterogeneity between individual cells and provide guidance for individualized treatment of early detection and diagnosis of diseases. One prominent problem in scRNA-seq data analysis is the prevalence of dropouts, caused by failures in amplification during the reverse-transcription step in

the RNA-seq experiment, which makes us to develop effective approaches for imputing the missed values. In this paper, we propose a novel imputation approach, referred to as the scLRTD, which use the low-rank tensor decomposition method (scLRTD) to impute scRNA-seq data. Furthermore, two sets of simulated data and a set of real data are used to carry out numerical experiments and compared with other six published methods. Error accuracy and clustering results demonstrate the effectiveness of our proposed method.

Paper ID: 58

Investigation of lipid metabolism dysregulation and the effects on immune microenvironment in pan-cancer using multiple omics data

Yang Hao[1,2], Daixi Li[1], Yong Xu[1,2], Jian Ouyang[2], Yongkun Wang[1,2], Yuqi Zhang[1,2], Baoguo Li[1], Lu Xie[2*] and Guangrong Qin[2*]*

[1]University of Shanghai for Science and Technology, China

[2]Shanghai Academy of Science and Technology, China

Lipid metabolism is a hallmark for tumor which contributes to tumorigenesis and progression, but the comparison of lipid metabolism among pan-cancer is not fully investigated. Through a systematically analysis of the multiple omics data from TCGA, we found that the most-widely altered lipid metabolism pathways in pan-cancer are fatty acid metabolism, arachidonic acid metabolism, cholesterol metabolism and PPAR signaling. Gene expression profiles of fatty acid metabolism shows commonalities across pan-cancer, while the alteration in arachidonic acid metabolism differs with tissue origin, suggesting tissue specific lipid metabolism features in different tumor types. An integrated analysis of gene expression, DNA methylation and mutation revealed factors that regulate gene expression, including the differentially methylated sites and mutations of the lipid genes, as well as mutation and differential expression of the up-stream transcription factors. Correlation analysis of the proportion of tumor immune microenvironment cells and lipid metabolism genes revealed immune-related differentially expressed lipid metabolic genes, suggesting the potential crosstalk between lipid metabolism and immune response. Genes related to lipid metabolism and immune response that are associated with poor prognosis were discovered, which may provide clues for tumor biomarkers or therapeutic targets.

Paper ID: 59

Computational prediction and functional analysis of arsenic binding proteins in human cells

Shichao Pang[1], Junchen Yang[1], Yilei Zhao[1], Yixue Li[2] and Jingfang Wang[1]*

[1]Shanghai Jiao Tong University, China

[2]Chinese Academy of Sciences, China

Arsenic has a broad anti-cancer ability against hematologic malignancies and solid tumors. To understand the mechanism of this broad anticancer ability, we developed a single mutation free energy profile for arsenic binding to identify arsenic binding proteins. Multiple validation provided an indication that our computational model could make successful prediction for arsenic binding proteins with accuracy. Additionally, we also applied this computational model to identify all the

potential arsenic binding proteins in the human genome. Functional analysis indicated that the arsenic binding proteins showed a wide range of biological functions, especially in the signaling transduction and related pathways. In signaling transduction pathways, arsenic could directly bind to the key factors (e.g., Notch receptors, Notch ligands, Wnt family proteins, TGF-beta, and their interacting proteins), and induce significant inhibition on their biological activities, further having crucial impact on the related signaling pathways. We hope that these findings could provide a genome-wide insight into the biological functions of arsenic, revealing a mechanism for the broad anticancer of arsenic.

Paper ID: 60

A novel joint gene set analysis framework improves identification of enriched pathways in cross disease transcriptomic analysis

Wenyi Qin[1], Xujun Wang[2] and Hui Lu[3]*

[1]Yale University, USA

[2]Shanghai Jiao Tong University, China

[3]University of Illinois at Chicago, USA

Motivation: Gene set enrichment analysis is a widely accepted expression analysis tool which aims at detecting coordinated expression change within a pre-defined gene sets rather than individual genes. The benefit of gene set analysis over individual differentially expressed (DE) gene analysis includes more reproducible and interpretable results and detecting small but consistent change among gene set which could not be detected by DE gene analysis. There have been many successful gene set analysis applications in human diseases. However, when the sample size of a disease study is small and no other public data sets of same disease are available, it will lead to lack of power to detect pathways of importance to the disease.

Results: We developed novel joint gene set analysis statistical frameworks which aims at improving the power of identifying enriched gene sets through integrating multiple similar disease data sets. Through comprehensive simulation studies, we demonstrated that our proposed frameworks obtained much better AUC scores than single data set analysis and another meta-analysis method in identification of enriched pathways. When applied to two real data sets, the proposed framework could retain the enriched gene sets identified by single data set analysis and exclusively obtain up to 200% more disease-related gene sets demonstrating the improved identification power through information shared between similar diseases. We expect that the proposed framework would enable researchers to better explore public data sets when the sample size of their study is limited.

Paper ID: 61

Screening drug combinations in disease-related molecular network

*Min Luo, Jianfeng Jiao and Ruiqi Wang**

Department of Mathematics, Shanghai University, Shanghai, China

Many complex diseases are generally caused by multiple factors. Drugs against multiple targets may combat those diseases. What's more, combination drugs may overcome many limitations of

single targets and achieve a more effective and safer control of the disease. Most of existing combination drugs are developed based on clinical experience or text-and-trial strategy, which are not only time consuming but also expensive. More novel, by exploiting high throughput data can identify effective drug combinations. However, systematic identification of multiple drug targets and their best intervention requires knowledge of the underlying disease network and calls for innovative computational methods that exploit the network structure and dynamics.

Paper ID: 62

DOS: A tool for predicting degree of specificity of monoclonal antibodies using sequences.

Anthony Mackitz Dzisoo and Jian Huang[*]
University of Electronic Science and Technology of China, Chengdu, China.

Monoclonal antibodies (mAbs) are one of the robust class of therapeutic proteins. Their stability, specificity, and high solubility allows the successful development and commercialization of antibody-based drugs. Though with these characteristics, mAbs projects are often suspended due to the low degree of specificity. This is one of the main reasons that causes the development of mAbs into drugs taking forever and expensive. The degree of specificity of monoclonal antibody can be quantified by the values of the following assays: Poly-Specificity Reagent (PSR), Affinity-Capture Self-Interaction Nanoparticle Spectroscopy (AC-SINS), Cross-Interaction Chromatography (CIC) and Clone Self-Interaction by Biolayer Interferometry (CSI-BLI). To save time and money, we developed a model called DOS which can predict degree of specificity based on single amino acid composition of mAbs sequences. It showed 98.9% accuracy, 98.1% sensitivity, 99.9% specificity, 0.97 Mathew Correlation Coefficient (MCC) , and 0.99 area under the receiver operating characteristic (ROC) curve (AUC) with the leave-one-out cross-validation. DOS is freely available at http://i.uestc.edu.cn/eli/cgi-bin/dos.pl.

Paper ID: 63

Revealing the Tipping Points in Infant Brain Development for Human and Chimpanzee by Gene Expression Data

Hui Tang, Ying Tang, Tao Zeng and Luonan Chen[*]
Chinese Academy of Sciences, Shanghai 200031, China

Postpartum developmental delay has been proposed as a possible biological mechanism of human evolution which contributes to many human-specific phenotypes, such as the increased brain size and the advanced human-specific cognitive traits. However, the biological processes and molecular functions during early brain development still remain poorly understood, especially in primates. In this paper, we extensively studied dorsolarteral prefrontal cortex expression data in human and chimpanzee to investigate the critical processes or biological events during early brain development at a molecular level based on dynamic network biomarker (DNB) theory [1-6]. We found that there are the tipping points around 3 months and 1 month corresponding to crucial periods in infant human and chimpanzee brain development, respectively. We also identified the DNBs and the differently expressed genes (DEGs) around the tipping points during the brain development of human and

chimpanzee respectively. In particular, we found that the human postnatal development and the expression changes are delayed 3 times relative to chimpanzee, and we also revealed that many common biological processes are involved in those periods for both human and chimpanzee. These findings all support that the maximal rates of brain growth are in those two critical periods for respective human and primates. In addition, our analytic results also indicate that human would further develop a number of advanced behavior functions around this tipping point (around 3 months), such as the ability of learning and memory. Meanwhile, chimpanzee acquires the ability of basic behaviors associated with survival at the tipping point. This work not only provides biological insights into the brain development at a molecular level but also opens a new way to study the criticality of nonlinear biological processes based on the observed omics data.

Paper ID: 64

Personalized critical variation of gut microbiota before the occurrence of Type I Diabetes

Lu Wang[1] and Luonan Chen[2*]*
[1]SIBCB, China
[2]Shanghai Institutes for Biological Sciences, China

Background / Aims:

As well-known, the gut microbiota is associated with many human complex diseases. The changes of microbiota community have been widely observed in disease individuals. However, the disorder of gut microbiota during disease occurrence is still unclear; especially such pre-disease or early-disease signal on individual gut microbiota requires systematical researches.

Subjects and Methods:

Different from conventional studies on differential average abundance of gut microbiota between normal and diseased samples, we investigated the variance of abundance of gut microbiota on consecutive samples from healthy to seroconversion for each individual because the change of abundance variance would be a critical signal of biological dynamical system. Single sample network (SSN) is a network reconstruction method based on the change caused from the new added sample to reference samples. And dynamical network biomarker (DNB) is a biomarker detection method to find the mostly distinguished subnet of markers before disease state transition. Here, we firstly used metagenomics data of total 15 individuals from three cohorts in public domain to build SSN networks, and each person in the cohorts would have multiple feces samples during more than 1 years. Then we used DNB to detect biomarkers from these SSNs.

Results:

6 individuals keep healthy, and 6 individuals occur seroconversion, and 3 individuals occur seroconversion along with final Type 1 Diabetes. Focused on the seroconversion, a key stage to T1D, the conventional analysis found several microbiota with changed abundance after the occurrence of seroconversion (Figure 1 The summary of microbiota abundance variance in dynamical network for different groups of samples). Meanwhile, our analysis further shows that many microbiotas actually have great abundance change before seroconversion (Figure 2 The personalized index of abundance variance change based on dynamic network biomarker theory), which can even be efficient features to classify the healthy and seroconversion individuals with about 80% accuracy.

Furthermore, these microbiota with critical abundance variance are also associated with T1D clinical antibody (Figure 3 The PCoA among microbiota and clinical indices).

Conclusions:

Dissimilar to common biomarkers like clinical antibody, the individual specific signatures, e.g. variance of gut microbiota abundance, would be alternative approach for personalized pre-disease or early-disease diagnosis.

Paper ID: 65

Biosystems Study of the Molecular Networks Underlying Hippocampal Aging Progression and Anti-aging Treatment in Mice

Jiao Wang[1], Qian Li[1], Yanyan Kong[2], Fangfang Zhou[1], Jie Li[1], Weihao Li[1], Kai Wang[3], Ting Wu[4], Yihui Guan[2], Jiang Xie[5] and Tieqiao Wen[1*]*

[1]Laboratory of molecular neural biology, School of life sciences, Shanghai University, China

[2]Position Emission Computed Tomography Center, Huashan Hospital, Fudan University, China

[3]Shanghai Key Laboratory of Molecular Andrology, Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, China

[4] Shanghai Stem Cell Group, China

[5] School of Computer Engineering and Science, Shanghai University, China

Aging progression is a process that an individual encounters as they become older, and usually results from a series of normal physiological changes over time. The hippocampus, which contributes to the loss of spatial and episodic memory and learning in older people, is closely related to the detrimental effects of aging at the morphological and molecular levels. However, age-related genetic changes in hippocampal molecular mechanisms are not yet well-established. To provide additional insight into the aging process, differentially-expressed genes of 3- versus 24- and 29-month old mice were re-analyzed. The results revealed that a large number of immune and inflammatory response-related genes were up-regulated in the aged hippocampus, and membrane receptor-associated genes were down-regulated. The down-regulation of transmembrane receptors may indicate the weaker perception of environmental exposure in older people, since many transmembrane proteins participate in signal transduction. In addition, molecular interaction analysis of the up-regulated immune genes indicated that the hub gene, Ywhae, may play essential roles in immune and inflammatory responses during aging progression, as well as during hippocampal development. Our biological experiments confirmed the conserved roles of Ywhae and its partners between human and mouse. Furthermore, comparison of microarray data between advanced-age mice treated with human umbilical cord blood plasma protein and the phosphate-buffered saline control showed that the genes that contribute to the revitalization of advanced-age mice are different from the genes induced by aging. These results implied that the revitalization of advanced-age mice is not a simple reverse process of normal aging progression. Our data assigned novel roles of genes during aging progression and provided further theoretic evidence for future studies exploring the underlying mechanisms of aging and anti-aging-related disease therapy.

Paper ID: 66

TriVote, a highly accurate OMIC biomarker detection algorithm

Fengfeng Zhou
Jilin University, China

Transcriptomic and methylomic patterns represent two major OMIC data sources impacted by both inheritable genetic information and environmental factors, and have been widely used as disease diagnosis and prognosis biomarkers. Modern transcriptomic and methylomic profiling technologies detect the status of tens of thousands or even millions of probing residues in the human genome, and introduce a major computational challenge for the existing feature selection algorithms. This study proposes a three-step feature selection algorithm, TriVote, to detect a subset of transcriptomic or methylomic residues with highly accurate binary classification performance. TriVote outperforms both filter and wrapper feature selection algorithms with both higher classification accuracy and smaller feature number on 17 transcriptomes and two methylomes. Biological functions of the methylome biomarkers detected by TriVote were discussed for their disease associations. An easy-to-use Python package is also released to facilitate the further applications.

Paper ID: 67

Integrated regulatory-metabolic network modeling and strain design based on the integrated model

*Fangzhou Shen, Renliang Sun, Jian Li and Zhuo Wang**
Shanghai Jiao Tong University, China

Gene regulatory and metabolic network models have been used successfully in many organisms, but inherent differences between them make networks difficult to integrate. Previous integration approaches required a pre-existed gene regulatory network from database or literatures, and only effective for prokaryotes, i.e. E.coli. We present a de-novo method IDREAM (Integrated Deduced REgulation And Metabolism) that combines statistically inferred regulatory network with metabolic network, and predict genetic interactions associated with gene mutants encoding transcription factors in Saccharomyces cerevisiae. The IDREAM model generated growth predictions significantly better correlated with experimentally measured growth than previous approach. Importantly, IDREAM's enhanced accuracy makes it possible to identify subtle synthetic growth defects. With experimental validation, these novel genetic interactions involving the pyruvate dehydrogenase complex suggested a new role for fatty acid-responsive factor Oaf1 in regulating acetyl-CoA production in glucose grown cells.

Computational strain optimization algorithms based on genome-scale metabolic models have increasingly been used to aid in overproducing products of interest. However, most of strain optimization algorithms only utilize a metabolic network alone, and few approaches can provide strategies involving transcriptional regulation. Based on the integrated IDREAM network, we developed a novel strain design algorithm, named OptRAM (Optimization of Regulatory And Metabolic Network), which can identify combinatorial optimization strategies including overexpression, knock down or knockout of both metabolic genes and transcriptional factors. OptRAM uses simulated annealing with a novel objective function, which can ensure a favorable

coupling between desired chemical and cell growth. The other advance we propose is a systematic evaluation metric of multiple solutions, by considering the essential genes, flux variation, and engineering manipulation cost. We applied OptRAM in succinate, 2,3-butanediol, and ethanol overproduction in yeast, which predicted high guaranteed target production rate compared with other methods and previous literature values. Moreover, the TFs and genes proposed to be altered are also predicted to have significant effects on the target overproduction above what can be achieved by metabolism manipulation alone. In conclusion, OptRAM can provide more comprehensive and effective strategies to assist biologists for particular metabolic engineering applications.

Paper ID: 69

De novo assembled individual genome doesn't show advantage against standard reference genome: a demonstration of Chinese Han Population

Zhibiao Mai, Wanting Liu, Wen Ding and Gong Zhang[*]
Jinan University, China

Millions of human genetic variations were found in the 1000 Genomes Project, including single nucleotide polymorphisms (SNPs) and structure variants (SVs). De novo assembly of individual genome is getting practical nowadays especially with the long-read single-molecule real-time (SMRT) sequencing for ethics-specific SVs. We took the complete genome sequence assembly of a Chinese Han individual HX1 as an example to evaluate whether de novo assembled individual genome has advantage in personalized medicine for the same ethnic. Surprisingly, all whole genome sequencing (WGS) data and whole exome sequencing (WES) data from various ethnics, including the Han, were mapped better to GRCh38 than to HX1. We used the Illumina short reads of HX1 to further correct HX1 genome, which improved the mapped rate to HX1. However, the mapped rate of the WGS/WES datasets to the corrected HX1 were still 1% lower than that of GRCh38. The high error rate of long reads from SMRT sequencing leads to misassembly: the mitochondrial genome was largely misassembled in HX1. We map the widely-used variant databases to HX1 genome to build HX1-specific variant databases for personalized medicine: 97.1% of dbSNP, 98.8% of ClinVar and 97.2% of COSMIC variants could be mapped to HX1. However, the HX1-specific regions almost did not contain any expressible genes, at least in the case of hepatocellular carcinoma cell lines and prostate cancer tissues originated from Chinese patients. Therefore these ethnics-specific regions may not play important role in precision medicine applications. Our results demonstrated that the de novo assembled individual genome doesn't show advantage against standard reference genome through various aspects and is not recommended in personalized medicine applications.

Paper ID: 70

Multifaceted stoichiometry control of bacterial operons revealed by data-independent acquisition mass spectrometry

Jing Zhao[1], *Hong Zhang*[1], *Bo Qin*[2], *Qingyu He*[1], *Rainer Nikolay*[2], *Christian Spahn*[2] *and Gong Zhang*[1*]
[1]*Institute of Life and Health Engineering, Jinan University, China*

[2]*Institut für Medizinische Physik und Biophysik, Charité-Universitätsmedizin, Germany*

More than half of the protein-coding genes in bacteria are organized in polycistronic operons composed of two or more genes. It remains under debate whether the operon organization maintains the stoichiometric expression of the genes within an operon. In this study, we performed a label-free data-independent acquisition hyper reaction monitoring mass-spectrometry (HRM-MS) experiment to quantify the Escherichia coli proteome in exponential phase and quantified 93.6% of the cytosolic proteins, covering 67.9% and 56.0% of the translating polycistronic operons in BW25113 and MG1655 strains, respectively. We found the shorter operons tend to be more tightly controlled for stoichiometry than longer operons, and the operons for metabolic pathways are less controlled for stoichiometry compared with operons for protein complexes, illustrating the multifaceted nature of the operon-wise regulation: the operon-wise unified transcriptional level and gene-specific translational level. This multi-level regulation benefits the host by optimizing the efficiency of the productivity of metabolic pathways and maintenance of different types of protein complexes.

Paper ID: 71

The ultrafast and accurate mapping algorithm FANSe3: mapping a human whole genome sequencing dataset within one hour

Gong Zhang[1,2], Yongjian Zhang[2] and Jingjie Jin[1]*
[1]*Jinan University, China*
[2]*Chi-Biotech Co. Ltd., China*

Aligning billions of reads generated from next-generation sequencing (NGS) to reference sequences, i.e. “mapping”, is the most time-consuming and computational intensive process among vast majority of the NGS applications. Fast, accurate and robust mapping algorithm is highly needed. To meet these demands, we developed FANSe3 mapping algorithm, which can map a 30x human whole genome sequencing (WGS) dataset within 1 hour, a 50x human whole exome sequencing (WES) dataset within 1 minute, and a typical mRNA-seq dataset within seconds, in a single server node without any hardware acceleration feature. Inherited from its predecessor FANSe2, the error rate can be kept as low as 10-9 in most cases with mathematical estimation, which is more robust than the Burrows-Wheeler Transform-based algorithms. The error allowance almost did not affect the result in clinical-relevant WGS to identify the driver somatic mutation and provided robust gene expression profiles regardless the parameter settings and the sequencers. This algorithm which designed for high-performance cloud computing infrastructures will break the bottleneck of NGS data analysis and promote various NGS applications faster and more robustly. The website of FANSe3: http://www.chi-biotech.com/fanse3/

Paper ID: 72

Pan-genome analyses of 24 Shewanella strains re-emphasize the diversification of their functions yet evolutionary dynamics of metal-reducing pathway

Kang Ning and Choafang Zhong*
Huazhong University of Science and Technology, China

Background: Shewanella strains are important dissimilatory metal-reducing bacteria which are widely distributed in diverse habitats. Despite efforts to genomically characterize Shewanella, knowledge of the molecular components, functional information and evolutionary patterns remain lacking, especially for their compatibility in the metal-reducing pathway. The increasing number of genome sequences of Shewanella strains offers a basis for pan-genome studies.
Results: A comparative pan-genome analysis was conducted to study genomic diversity and evolutionary relationships among 24 Shewanella strains. Results revealed an open pan-genome of 13,406 non-redundant genes and a core-genome of 1,878 non-redundant genes. Selective pressure acted on the invariant members of core genome, in which purifying selection drove evolution in the housekeeping mechanisms. Shewanella strains exhibited extensive genome variability, with high levels of gene gain and loss during the evolution, which affected variable gene sets and facilitated the rapid evolution. Additionally, genes related to metal reduction were diversely distributed in Shewanella strains and evolved under purifying selection, which highlighted the basic conserved functionality and specificity of respiratory systems.
Conclusion: The diversity of genes present in the accessory and specific genomes of Shewanella strains indicates that each strain has used different strategies to adapt diverse environments. Horizontal gene transfer is an important evolutionary force in shaping Shewanella genomes. Purifying selection has played an important role in stability of the core-genome and also drives evolution in mtr-omc cluster of different Shewanella strains.

Paper ID: 73

The carcinogenic molecules were revealed by novel integrative analysis

Wanting Liu[*] and Gong Zhang[*]
Jinan University, China

The multi-center expression omics data provides new approaches for researching carcinogenic molecules, however, data preference still exists in the approach which are caused by the differences of platforms, reagents and instruments. These differences will interfere in common meta-analysis methods to cover the meaningful molecules when analyze mass data. To deal with these issues, our strategy will apply the novel integrative analysis which has robustness, high fault tolerances and strong convergences to solve out the system deviation. So, the key differential expressed molecules can be retrieved by using this strategy to integrative analysis multi-center omics data.
The integrative strategy is set up based on the rank method to eliminate the central system deviation and integrated the multicenter cancer/precancerous microarray datasets to fight the shortage of existing integrative methods. Using this strategy we were successfully selected 12 biomarker signature of melanoma, and 4 of them were the first found that they were relevant to melanoma. We also applied this strategy to 330 microarray datasets of five multi-center lung adenocarcinoma studies and retrieved 200 biomarker signatures. Based on this result, we found 3 novel key molecules of lung adenocarcinoma with strong survival correlation. Next, we explored multicenter breast cancer microarray datasets, and selected 64 key molecules, 13 of these molecules were first reported that they are relevant with breast cancer. We already randomly validated four molecules and determined they all related with overall survival time of breast cancer.

The novel integrative strategy we developed can avoid the system deviation of multiple datasets, thereby improving the number of sample datasets from previous studies, it is benefit for detecting the universal key molecules. In multiple cancer studies, we all found the unreported key differentially molecules with verifiable molecular functions and prognostic significances. This shows that the novel integrative strategy can retrieve the new and valuable key molecules for cancer research and clinical precise application.

Paper ID: 74

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

Jing Ge[1], Gaopeng Li[2], Shuxian Li[3], Huiyong Yin[4] and Luonan Chen[5*]*

[1]Institute of Biochemistry and Cell Biology, SIBS, CAS, China

[2]Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China

[3]School of Life Science and Technology, ShanghaiTech University, China

[4]Institute for Nutritional Sciences, SIBS, CAS, China

[5]Osaka Sangyo University, China

Atherosclerosis is one of the major factors causing cardiovascular diseases. The onset and progression of atherosclerosis is a complex and nonlinear process, which involves complicated dynamic regulations among biomolecular networks. However, the major metabolic networks that regulate the homeostasis of blood vessel remain poorly defined. Our project combines omics data and theoretical analysis to establish a theoretical model for early prediction of the dysregulation of blood vessel homeostasis and atherosclerotic lesion formation from the viewpoint of systems biology. To be specific, first we used the LDLR-/- mouse model feeding western diet to mimic the onset and progression of atherosclerosis. The pathological characteristics of aortic roots by Oil-red-O and H&E staining showed that the atherosclerotic lesion rapidly increased and deteriorated after 8 weeks. Next, we integrated high throughput data including different-stage data of RNA-Seq and lipidomics to discover 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 tipping point in the progression of atherosclerosis at the genetic level and discovered a group of DNB genes which appear to play driving roles in the progression of this disease. Moreover, lipidomics of plasma further verified the critical point we identified at the genetic level, which may provide a new insight for noninvasive prediction of atherosclerosis. At last, we attempted to validate the theoretical results with biological experiments. In summary, we aim to develop a prediction model to identify the early signals of the critical transition in atherosclerosis, which may help for the early detection and prevention of atherosclerosis.

Paper ID: 75

Systems Biology Theory Resolution of a Controversy in Pancreatic Beta Cell Regeneration

Haoran Cai[1], Runtan Cheng[2], Xiaomei Zhu[1] and Ping Ao[1]*

[1]Shanghai University, China

[2]*Shanghai Jiao Tong University, China*

Researchers have long debated regarding a controversial question: whether new pancreatic β-cells arise via differentiation of precursor cells or from pre-existing β-cells. The concept that the formation of new endocrine cells is from the stem or precursor cells has been active since the late 19th-century. Recent experiments suggested that multipotent precursors from adult mouse pancreas do exist. However, such finding has been challenged by two main lines of research. The first involves that Neurog3 (marks precursors of islet) is undetectable in the adult pancreas. Another evidence is related with β-cell self-replication. It has been recognized that adult β-cell retains a limited capacity for proliferation. More surprisingly, seminal lineage-tracing study found that the fraction of labeled beta-cells remained unchanged over a one-year chase period, suggesting that self-replication is the exclusively major source of β-cell expansion, which is further amplified by subsequent confirmatory reports. On the other hand, more possibility was found for this type of β-cells labeled that the cells can be multipotent precursors. Indeed, the observations showed that the multipotent precursors of β-cells did express insulin in vivo. Yet another group identified non-insulin-expressing cells in islets that could give rise to beta-cells. Do multipotent precursors exist within pancreas?

Here we use a systems biology approach, the endogenous network theory, to construct an endogenous molecular cellular network for pancreas, reproducing the core features of multiple cellular phenotypes from robust stable states. We sought to analyzed the natural mechanisms by which beta cells are formed during adult life in a network level, where the seemingly controversial observations of β-cell expansion can be all integrated into a single model. Hence a clear understanding of the origin of the debate from a systems biology theory has been obtained. A brief introduction to endogenous network theory and some of its related recent progress, such as its ability to predict cell states and lineage, will be presented, too.

References:

Beyond cancer genes: colorectal cancer as robust intrinsic states formed by molecular interactions
Yuan et al., Open Biol. 2017, 7: 170169.
http://rsob.royalsocietypublishing.org/content/royopenbio/7/11/170169.full.pdf
Decoding early myelopoiesis from dynamics of core endogenous network.
Su et al., Science China Life Sciences, 2017, 60, 627-646
http://engine.scichina.com/publisher/scp/journal/SCLS/60/6/10.1007/s11427-017-9059-y?slug=full
Cancer as Robust Intrinsic State of Endogenous Molecular-Cellular Network Shaped by Evolution.
Ao, Galas, Hood, Zhu. Medical Hypothesis, 2008, 70, 678–684.
http://dx.doi.org/10.1016/j.mehy.2007.03.043
https://www.ncbi.nlm.nih.gov/pubmed/17766049

Paper ID: 76

A hybrid deep learning method integrating CNN and BLSTM in predicting protein phosphorylation sites

Haixia Long[1], Bo Liao[1], Jialiang Yang[1*], Yuhua Yao[1] and Luonan Chen[2]*
[1]*HaiNan Normal University, China*

[2]*Chinese Academy of Sciences, China*

Protein phosphorylation is one of the most important post-translational modifications (PTMs) occurring at serine (S), threonine (T), and tyrosine (Y) amino acid. It plays critical roles in many biological processes including signal transduction, cell cycle, and cancer development. With the development of novel high-throughput sequencing technologies, there are huge amount of protein sequences being generated and stored in databases. It is of great importance in both basic research and drug development to quickly and accurately predict if an uncharacterized protein sequence containing many residues of S, T, or Y can be phosphorylated. In order to address the problem, a novel hybrid convolutional and bi-directional long short-term memory recurrent neural network model (CNN+BLSTM) is proposed for predicting phosphorylation sites in proteins. The model contains a list of layers that transform the input data into an output class, in which the convolution layer captures higher-level abstraction features of amino acid, while the recurrent layer captures long-term dependencies between amino acids to improve predictions. We applied our model together with two canonical methods namely CNN and random forest to a published benchmark data. A 5-fold cross validation process indicated that CNN+BLSTM outperforms the two competitors in various evaluation metrics like the area under receiver operating characteristic and precision-recall curves, the Matthews correlation coefficient, F-measure, accuracy, and so on. CNN+BLSTM is promising in identifying potential protein phosphorylation for further experimental validation.

Paper ID: 77

Systematic survey and prediction reveal widespread context-dependent activities of RNA binding proteins in splicing regulation

Yue Hu, Miaowei Mao and Zefeng Wang[*]

CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, China

Alternative splicing (AS) is generally regulated by splicing factors that specifically bind to cis-elements in pre-mRNAs. Human genome encodes ~1500 RNA binding proteins (RBPs) that potentially regulate AS, yet their functional domains and activities remain largely unknown. To explore their potential activities, we fused the putative functional domains of various RBPs to a sequence-specific RNA-binding domain, and systemically analyzed how these engineered factors affect splicing of reporter genes. We discovered that ~80% of low complexity domains in endogenous RBPs displayed distinct context-dependent activities in regulating splicing, indicating that AS is under more extensive regulation by a variety of RBPs than previously expected. We developed a machine learning approach using the experimental results as a training dataset to classify and predict the activities of RBPs based on their sequence compositions, and further validated this model using endogenous RBPs with unknown functional domains and synthetic polypeptides. These results represent the first systematic inspection, modeling, prediction and validation of how RBP sequences affect their activities in controlling splicing, paving the way for de novo engineering of artificial splicing factors.

Paper ID: 79

RBM10 functions as a tumor suppressor in lung cancer by mediating alternative splicing of key target genes

Sirui Zhang[1], Yongbo Wang[2] and Zefeng Wang[1*]*
[1]CAS-MPG Partner Institute for Computational Biology, China
[2]Fudan university, China

Alternative splicing (AS) influences the expression of most eukaryotic genes and dysregulation of AS is one of the molecular hallmarks of cancer. RBM10 is a splicing factor-encoding gene that is frequently mutated in in lung adenocarcinoma (LUAD), but its functions and underlying molecular mechanisms in LUAD pathogenesis remains elusive. Here we characterized RBM10 mutations in East Asian lung adenocarcinoma (LUAD) patients and compared with those in Western LUAD patients reported by The Cancer Genome Atlas (TCGA). We found that the majority of RBM10 mutations are loss-of-function mutations in LUAD patients from both populations, despite their dramatically distinct patterns of oncogenic mutations. Co-mutation analysis revealed that RBM10 mutations co-occurred with oncogenic mutations. To gain molecular insights into RBM10 functions in LUAD, we conducted RNA-Seq experiments (1) following RBM10 knockdown (KD) in lung epithelial cells and overexpression (OE) in LUAD cells, respectively, and (2) in LUAD tissues with and without RBM10 mutation as well as their matched adjacent non-tumor tissues. In addition, we performed integrative analysis of mutation, expression and clinical data of LUAD samples deposited in TCGA. Results from these experiments and data analyses suggest that RBM10 exerts suppressive functions in LUAD by regulating splicing of key target genes. Strikingly, several RBM10-regulated RNA splicing events in LUADs are significantly associated with patient survival. Overall, this study provides evidence for tumor suppressor functions and clinical significance of RBM10 in lung cancer and mechanistic insights into cancer-related splicing dysregulation .

Paper ID: 80

Subnetwork identification and chemical modulation for neural regeneration. A study combining network guided forest and heat diffusion model

Hui Wang[1], Gang Wang[1], Lida Zhu[1], Xuan Xu[1], Bo Diao [2] and Hong-Yu Zhang[1]*
[1]Huazhong Agricultural University, China
[2]Department of Clinical Experiment, Wuhan General Hospital of Guangzhou Command, China

Background: The induction of neural regeneration is vital to the repair of spinal cord injury (SCI). While compared with peripheral nervous system (PNS), the regenerative capacity of the central nervous system (CNS) is extremely limited. This indicates that modulating the molecular pathways underlying PNS repair may lead to the discovery of potential treatment for CNS injury.
Methods: Based on the gene expression profiles of dorsal root ganglion (DRG) after a sciatic nerve injury, we utilized network guided forest (NGF) to rank genes in terms of their capacity of distinguishing injured DRG from sham-operated controls. Gene importance scores deriving from NGF were used as initial heat in a heat diffusion model (HotNet2) to infer the subnetworks underlying neural regeneration in the DRG. After potential regulators of the subnetworks were

found through Connectivity Map (cMap), candidate compounds were experimentally evaluated for their capacity to regenerate the damaged neurons.
Results: Gene ontology analysis of the subnetworks revealed ubiquinone biosynthetic process is crucial for neural regeneration. Moreover, almost half of the genes in these subnetworks are found to be related to neural regeneration via text mining. After screening compounds that are likely to modulate gene expressions of the subnetworks, three compounds were selected for the experiment. Of them, trichostatin A, a histone deacetylase inhibitor, was validated to enhance neurite outgrowth in vivo via an optic nerve crush mouse model.
Conclusions: Our study identified subnetworks underlying neural regeneration, and validated a compound can promote neurite outgrowth by modulating these subnetworks. This work also suggests an alternative approach for drug repositioning that can be easily extended to other disease phenotypes.

Paper ID: 81

Inference of differentiation time for single cell transcriptomes using cell population reference data

Na Sun[1], Xiaoming Yu[2], Fang Li[1], Denghui Liu[1], Shengbao Suo[1], Weiyang Chen[1], Shirui Chen[3], Lu Song[3], Christopher D. Green[1], Joseph McDermott[1], Qin Shen[2], Naihe Jing[3] and Jing-Dong J. Han[1]*
[1]Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, China
[2]Tsinghua-Peking Center for Life Sciences, Tsinghua University, China
[3]Shanghai Institutes for Biological Sciences, China

Single-cell RNA sequencing (scRNA-seq) is a powerful method for dissecting intercellular heterogeneity during development. Conventional trajectory analysis provides only a pseudotime of development, and often discards cell cycle events as confounding factors. Here using matched cell population RNA-seq (cpRNA-seq) as a reference, we developed an "iCpSc" package for integrative analysis of cpRNA-seq and scRNA-seq data. By generating a computational model for reference "biological differentiation time" using cell population data and applying it to single cell data, we unbiasedly associated cell-cycle checkpoints to the internal molecular timer of single cells. Through inferring a network flow from cpRNA-seq to scRNA-seq data, we predicted a role of M phase in controlling the speed of neural differentiation of mouse embryonic stem cells, and validated it through gene knockout experiments. By linking temporally matched cpRNA-seq and scRNA-seq data, our approach provides an effective and unbiased approach for identifying developmental trajectory and timing related regulatory events.

Paper ID: 82

Identifying the patterns of double-stranded break sites during meiosis homologous recombination

Qiu Wang
Shanghai Institute of Biochemistry and Cell Biology, CAS, China

The ratio of species evolution has been greatly facilitated since the appearance of meiosis. There

are two major differences between meiosis and mitosis. One is homologous recombination which makes exchange between homologous chromosomes during meiosis prophase I. And another one is the consequence which produces haploid rather than diploid. The happen of programmed double-stranded breaks(DSBs) is seen as the initiation of homologous recombination. On the basis oof large datasets, we found that DSBs were not randomly distributed on the genome but preferred located in the intergenic regions. What's more, there was high correlation between H3K4me3 modification and DSBs. Scientists want to reveal how the DSBs sites are chosen during meiosis through the distribution of H3K4me3, which fail to perfectly match the DSBs distribution. As all of the messages of cell should be hidden in the DNA sequence, we tent to decode the mechanism of DSBs site distribution pattern by using HMM or other biostatistics methods.

Paper ID: 83

Building a Knowledgebase for Precision Medicine

Lei Liu
Fudan University, China

With the development of science and technology has entered intelligent era, Biomedicine will become more intelligent and precise. The trend of precision medicine study is the integration of various new technologies including big data technology, omics technology, bioinformatics, and so on. The research in life sciences has gradually turned into the mode of data-driven discovery. The key step in the process is the understanding and annotating the biomedical big data. We believe that the integrated analysis of clinical and bioinformatics data can make breakthrough in precision medicine. Precision medicine will revolutionize the medical practice, allow doctors make personalized and precise diagnosis, treatment and prevention plan based on patients' personal genomic data, life style and other specific factors.

Paper ID: 101

EEG based Intelligent robot control

Li Kun
School of Life Sciences, Guizhou Normal University

Electroencephalogram (EEG) contains some special physiological information. Based on the brain computer interface P300-speller, by using various classifications and pattern recognition techniques, such as wavelet transform, independent component analysis, principal component analysis, support vector machine and so on, we separated the special component P300 from EEG and judged the direction of movement to control the intelligent robot. In the future, the core technology of the EEG control robot is expected to be used in the medical rehabilitation industry. It brings hope to the high paraplegic patients. Further more, this technology can be applied to the military industry. Intelligent robot controlled by EEG will definitely speed up the development of artificial intelligence.

Paper ID: 102
基因测序数据处理的高性能计算挑战
朱红
浪潮集团 高性能应用支持专家

当前，以二代测序和三代测序为主流的高通量基因测序技术迅速普及。和二代测序不同的是，一方面三代测序的读长更长，但是另一方面测序的错误率也更高。这使得三代测序的数据比较适合做基因组组装，但是在进行组装前又往往面临着较为复杂的测序错误校正问题。后者导致三代测序组装的计算量极大，也对高性能计算服务器的配置带来了挑战。我们测试了以 FALCON 为代表的三代测序软件，对三代测序软件的优化做了一些尝试，并从将计算的角度给出了一些集群配置和使用的建议。

Paper ID: 103
Illumina 生物信息学整体解决方案
唐顺江
Illumina 大中华区生物信息平台总监

介绍 Illumina 在生物信息学领域端到端的整体解决方案。从实验室文库制备跟踪、测序过程监控到数据分析、生物学信息解读等各个流程环节，Illumina 均能够为客户提供了相应的生物信息学解决方案，其中包括：BaseSpace 平台、基于 FPGA 芯片技术的 Dragen 系列分析应用套件以及其他相配套的生物信息学工具等。

Paper ID: 104
Distinct endophytic communities associate with different plants for adaptation to karst environments
*Fei Li, Xiaohong He, Yuanyuan Sun, Ximin Zhang, Xiaoxin Tang, Yuke Li, Yin Yi**
The Key Laboratory of biodiversity conservation in Karst mountain area of Southwest of china, Forestry Ministry, Guizhou Normal University
Key Laboratory of Plant Physiology and Developmental Regulation, Guizhou Normal University
School of Life Sciences, Guizhou Normal University, Guiyang, Guizhou, China

Plants live in symbiosis with endophytic microbial communities which play vital roles in the physiological and pathological processes of the host plant. Since endophytes are in general horizontally transferred, the composition of endophytic microbial communities is affected by the geography and the ecosystem of the area where the host plant is grown. Studies have also shown that the characteristics of the host plant can also be a determinant factor defining the endophytic community structures. Our previous investigations indicated that there are significant differences in the calcium content of dominant plants growing in karst areas, suggesting that these plants may adopt different strategies for the absorption and utilization of calcium in the karst ecological environment. However, studies regarding the effects of plants growing in karst regions on their symbiont microbial communities are scarce and whether these plants adapt to calcium stress through mechanisms driven by the endophytic communities is unknown. The present study is, to the best of our knowledge, the first systematic survey of endophytic communities of dominant plants in the

karst ecosystem. Soil and plant materials were collected and after sequencing of the 16s RNA, the diversity and abundance of the endophytic community structures in leaves were examined.

Paper ID: 105

Deep learning helps optimize CRISPR guide RNA design

Guohui Chuai[1,2], Hanhui Ma[5], Jifang Yan[1,2], Ming Chen[4], Nanfang Hong[1,2], Dongyu Xue[1,2], Chi Zhou[1,2], Chenyu Zhu[1,2], Ke Chen[1,2], Bin Duan[1,2], Feng Gu[6], Sheng Qu[1,2], Deshuang Huang[3], Jia Wei [4*]and Qi Liu[1,2*]*

[1]Department of Endocrinology & Metabolism, Shanghai Tenth People's Hospital, Tongji University, Shanghai 20009, China

[2]Bioinformatics Department, School of Life Science and Technology, Tongji University, Shanghai 20009, China

[3]Machine learning & Systems Biology Lab, School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

[4]R&D Information, Innovation Center China, AstraZeneca, 199 Liangjing Road, Shanghai 201203, China

[5]School of Life Science and Technology, ShanghaiTech University, Shanghai, China

[6]State Key Laboratory Cultivation Base and Key Laboratory of Vision Science, Ministry of Health and Zhejiang Provincial Key Laboratory of Ophthalmology and Optometry, School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China

A major challenge for effective application of CRISPR systems is to accurately predict the single guide RNA (sgRNA) on-target knockout efficacy and off-target profile, which would facilitate the optimized design of sgRNAs with high sensitivity and specificity.

Here we present DeepCRISPR, a comprehensive computational platform to unify sgRNA on-target and off-target site prediction into one framework with deep learning, surpassing available state-of-the-art in silico tools.

In addition, DeepCRISPR fully automates the identification of sequence and epigenetic features that may affect sgRNA knockout efficacy in a data-driven manner

illumina®

BaseSpace® Informatics Suite

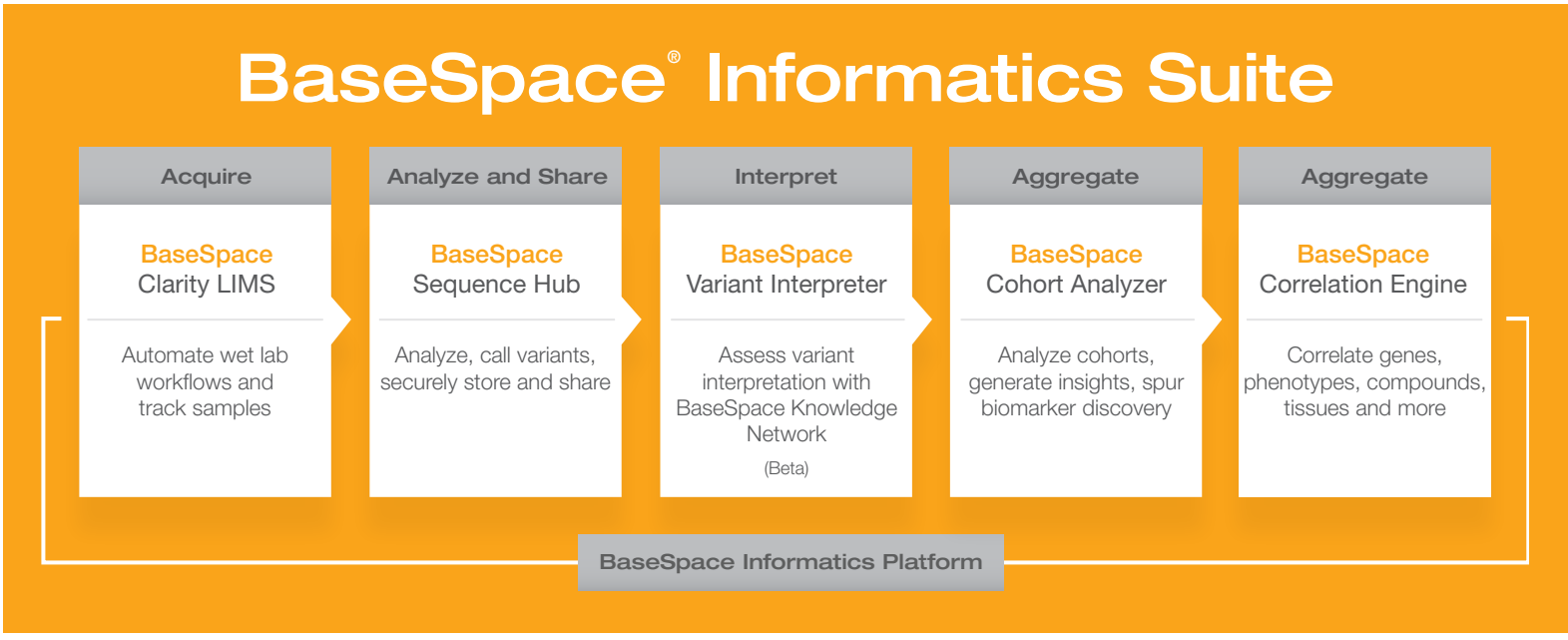
The shortest distance between sample and answer

Advances in genomic technologies promise to shed light on disease and how we diagnose and treat it. Yet the process for progressing from biological sample to meaningful answer is disconnected, arduous, and time-consuming. Disparate data systems and the need to assemble a diverse set of software applications create a highly manual and error-prone process that can take weeks to complete. To address these challenges, we proudly introduce BaseSpace® Suite, an integrated informatics product portfolio that unifies Illumina software products to deliver a comprehensive solution for genomic data analysis.

Simplify and expedite your genomic workflows with integrated informatics

BaseSpace Suite removes common informatics barriers by combining key functionality into one, easy-to-use, integrated solution. BaseSpace Suite is a comprehensive, cloud-based portfolio of informatics solutions built upon a software platform that provides common capabilities across the suite. Unifying key functionality, such as laboratory information management, data storage, analysis, and interpretation, means that you can more quickly deliver high-quality genomic information and apply the results to research and translational applications.

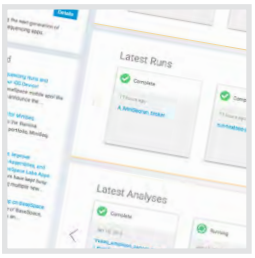
BaseSpace Suite consists of the following solutions that you can use together, separately or with the existing solutions in your lab.



illumina®

Track and manage samples and optimize wet lab workflows

BaseSpace Clarity LIMS (formerly Clarity LIMS) is a laboratory information management system that includes preconfigured workflows, support for regulatory compliance, flexibility to adapt to new workflows, and automation to help your lab scale when sample volumes increase.

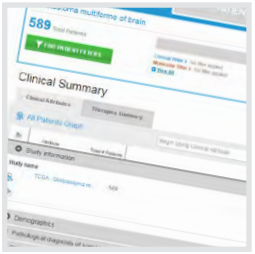


Analyze, store, and share genomic data

BaseSpace Sequence Hub (formerly BaseSpace) provides scalable and secure storage and analysis of genomic data. You can also remotely monitor sequencing runs, as well as share data to collaborate with colleagues.

Assess variant significance

BaseSpace Variant Interpreter (Beta) enables you to perform rapid annotation, filtering, interpretation, and reporting of genomic data. Accelerate variant interpretation with **BaseSpace Knowledge Network**, an integrated knowledge base containing genotype-phenotype associations.



Increase understanding of clinically and biologically significant markers

BaseSpace Cohort Analyzer (formerly NextBio Clinical) enables you to automatically aggregate and analyze subjects with genomics data in a few clicks. Review cohorts for marker frequencies, response, and outcomes and share data for biomarker discovery, translational research, and clinical trials.

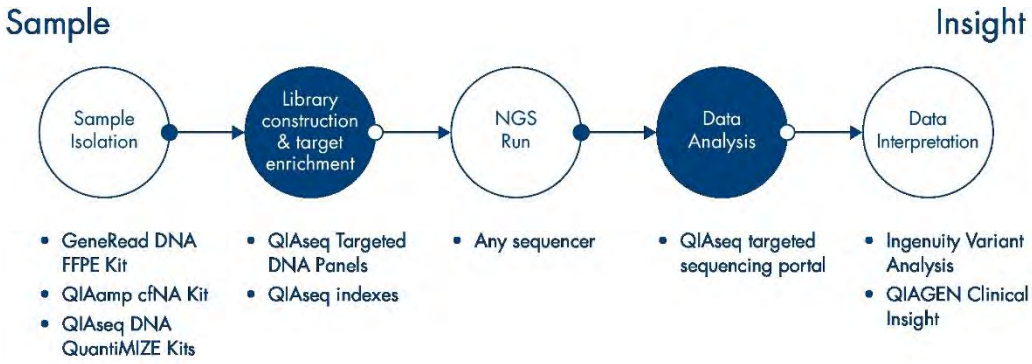
Use data-driven answers to understand genes, variants and signatures

BaseSpace Correlation Engine (formerly NextBio Research) enables you to instantly mine over 20,000 studies to get data-driven answers for genes, variants and signatures.

For more information, contact your Illumina sales representative.
www.illumina.com/informatics

Pub. No. 970-2016-019 Current as of 28 September 2016

QIAGEN 作为 Sample to Insight 全球领导者，在样本制备、文库构建和数据分析及解读方面精耕细作，开发和整合了众多完整解决方案。



在样本制备方面，提供具有诸多专利技术的产品：

- **PAXgene ccfDNA tube**：专利材质，非交联原理的稳定试剂，最大程度保证红细胞稳定性，避免基因组 DNA 污染，同时确保游离核酸的高保真性。
- **GeneRead DNA FFPE Kit**: 人为修复 FFPE 样本普遍存在的 C->T 随机突变，提高后期结果的准确性，仅需一片切片组织就能进行高品质 DNA 纯化，更可利用 QIAcube 或 QIAsymphony 进行自动化纯化。
- **QIAamp DNA Microbiome Kit**：采用分步骤裂解的原理，该试剂盒能最大程度消除高达 90%以上来自宿主 DNA 对测序数据的干扰。

在文库构建方面，无论是针对全基因组、扩增子、小 RNA、转录组以及最近非常火热的免疫组库、DNA 甲基化文库构建, QIAGEN 均推出了非常有针对性的解决方案。此外，基于**分子条形码 UMI** 和**单端特异性引物延伸（SPE）扩增技术**的 QIAseq Panel 集 Panel 富集和文库构建与一体，最大程度降低 PCR 扩增过程中引入的 PCR 错误、GC 偏好性和 duplication Rates。

- **QIAseq FX Library Prep Kit**: 基于全新酶解法的一体化文库试剂，对于不同 GC 含量的区域打断均一性好、 文库多样性好、duplicates 率低。
- **QIAseq™ miRNA Library Kit**: 基于 UMI 技术，仅需 1ng 总 RNA，7 个小时完成小 RNA 文库构建，无需割胶回收。
- **QIAseq cfDNA All-in-One Kit**：一体化游离核酸方案，涵盖从游离核酸纯化到 cfDNA 文库构建，更简单更省钱。
- **QIAseq Immune Repertoire RNA Library Kit**：免疫组库研究方案，基于 UMI 技术，对于 TCR α，β，γ 及 σ 全部亚基的完整可变区域 PCR 扩增，确保扩增结果均一性显著提升。

在数据分析及解读方面，QIAGEN 一直致力于为更多生物学家和生物信息学家提供一个直观、易用并且精准的桌面化分析方案。对变异检测算法进行优化和开发，最大程度降低假阳性位点；构

建包含 1,300 万生物学证据的数据库并引入行业的标准指南（ACMG/ASCO/AMP/CAP）对变异位点进行最精准的解读和分类。

- **Biomedical Genomics Workbench**: 基于 Windows 桌面系统，支持从原始数据（FASTQ/BAM）到变异位点及差异表达基因的分析；
- **Ingenuity Variant Analysis**：基于 QIAGEN Knowledge Base 对变异位点进行注释和筛选，快速锁定与表型相关的致病变异位点；
- **Ingenuity Pathway Analysis**：基于 QIAGEN Knowledge Base 深入挖掘差异表达基因参与的生物学通路和对应的生物学功能，阐明分子之间的调控机理；
- **HGMD**: 基于高质量的文献结果，人工构建遗传性疾病变异位点金标准数据库；
- **QCI-I:** 记忆 QIAGEN Knowledge Base 及行业标准指南进行变异位点的临床意义解读
- **OmicSoft**: 通过整合国际大型疾病研究项目相关数据库（包括 TCGA、ICGC、GTEx 等），简单、快速挖掘大队列基因组学信息；

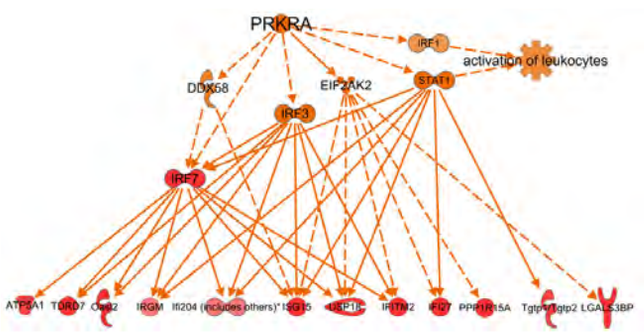
LUCIDUS BIO
路思达生物

简介 BRIEF INTRODUCTION

北京路思达生物信息科技有限公司成立于 2010 年，致力于为广大科研工作者提供最优质的生物信息学分析软件、数据库和数据分析课程，为您的科研之路提供便利。主要生物信息软件和数据库产品有：通路分析软件（Ingenuity Pathway Analysis, IPA）、人类基因突变数据库（Human Gene Mutation Database, HGMD）、变异分析软件（Ingenuity Variant Analysis, IVA），以及 CLC Bio 系列的二代测序数据分析软件 CLC Genomics Workbench、Biomedical Genomics Workbench，这些软件和数据库也常用在精准医学数据分析中。

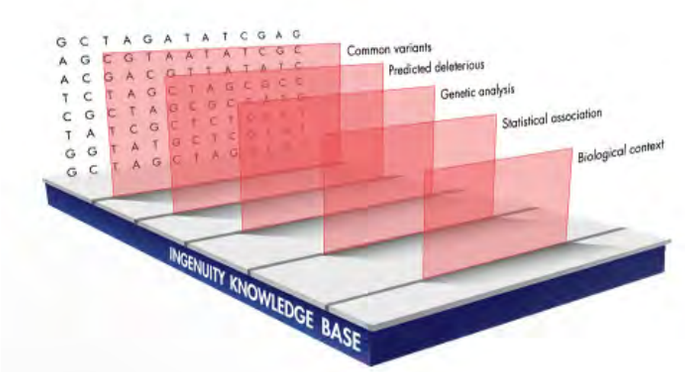
1. INGENUITY Pathway Analysis – IPA

基于人工矫正的知识库 QIAGEN Knowledge Base，可实现转录组、代谢组、蛋白质组、磷酸化蛋白质组等组学数据的经典通路分析、上游调控因子分析、下游疾病和功能分析、调控效应网络分析等；也可实现 microRNA 和 mRNA 组学数据的联合分析；图形美观，可直接用于发表，助您轻松实现文献调研、实验设计和文章发表。



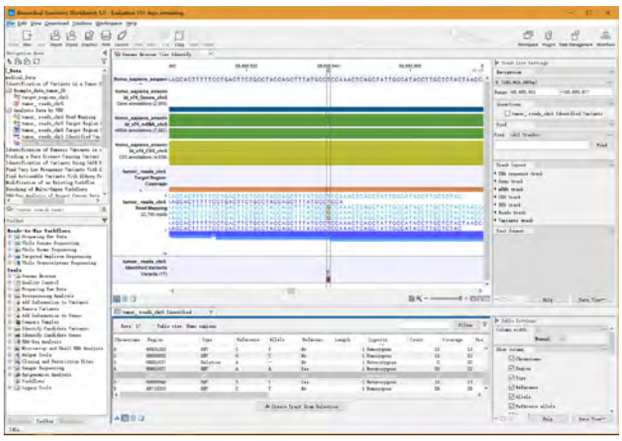
2. INGENUITY Variant Analysis – IVA

基于人工矫正的知识库 QIAGEN Knowledge Base，快速从人类测序数据中鉴定出具有说服力的致病变异，仅需数小时即可完成数百例样本的分析。IVA 提供的过滤项有：Biological Context、Cancer Drive Variants、Common Variants、Confidence、Custom Annotation、Genetic Analysis、Pharmacogenetics、Phenotype-Driven Ranking、Physical Location、Predicted Deleterious、Statistical Association、User-Defined Variants。



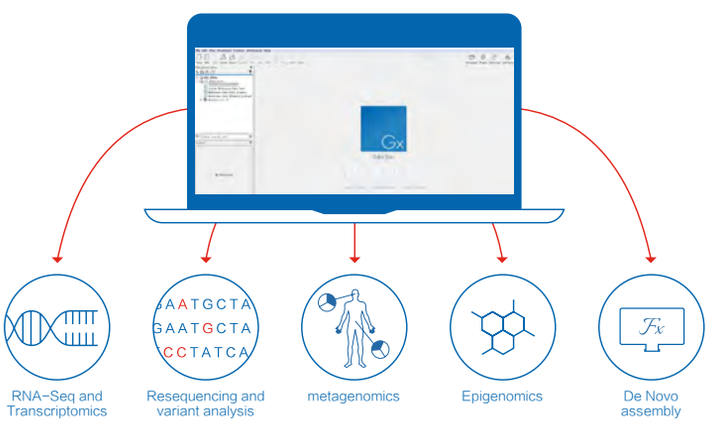
LUCIDUS BIO

3. Biomedical Genomics Workbench – BGWB



专注于生物医学领域内的高通量测序数据分析：具有高灵敏度的生殖细胞系变异检出和低频变异检出功能；支持 CNV 检出；探索新的生物标志物；已预置各种常用的 Workflow，可实现全基因组、全外显子组、全转录组和靶标测序数据的一键式分析；一键下载常用参考数据库；支持表观基因组学数据分析。

4. CLC Genomics Workbench – CLC GWB



可视化的高通量测序数据分析软件，适用所有主流测序平台产出的测序数据，囊括各种主流数据分析功能，如宏基因组分析、重测序分析、转录组分析、小 RNA 分析、表观分析和从头拼接等，Workflow 工具可实现大批量数据的一键式分析，支持自定义安装各类插件。

5. Human Gene Mutation Database – HGMD

人类基因突变数据库，致力于收录已发表的与人类遗传病相关的核基因突变，是解析遗传病的金标准。所收录的突变依据致病相关性分为：致病突变（DM）、可能致病的突变（DM?）、功能多态性（FP）、有功能影响的并与疾病有相关性的突变（DFP）和有疾病相关性的突变（DP）。

北京路思达生物信息科技有限公司
Beijing Lucidus Bioinformation Technology Co.,Ltd

北京市海淀区上地农大南路 33 号中电华软大厦 6 层
6th Floor, China Electronics Huaruan Mansion, No. 33 Nongda South Road,Shangdi, Haidian, Beijing

Website: www.lucidusbio.com Tel: 010-82696760 Email: Info@lucidusbio.com

Inspur Group Introduction

-Standard Version-

Inspur group is China's leading cloud computing, big data service provider with four listed companies: Inspur Information, Inspur Software, Inspur International and Inspur Huaguang, covering four industry groups of cloud data center, cloud services & big data, smart city, and smart enterprise and providing comprehensive solutions in IaaS, PaaS, and SaaS.

With high-end servers, mass storage, cloud operating system and information security technology, Inspur helps customers to build leading platforms underpinned by cloud computing infrastructure; and based on its information software for governments, enterprises, and industries, terminal products, and solutions, Inspur fully supports governments, enterprises, and industries in cloud construction. Inspur has shipped its IT products and services to more than 100 countries and regions around the world.

Inspur ranks top 2 in China’s IT industry with its comprehensive strength, and top one among all China’s domestic software vendors. Inspur is the 3rd largest server provider in the world and No.1 in China; No.1 storage seller in China for 13 consecutive years, No.1 ERP provider in China for 14 years. Inspur Cloud has the largest share in China’s government cloud market and

Inspur Tianyuan Data ranks No.1 in China’s big data market. Inspur is one of the eight National Secure and Reliable Computer Information System Integration Key Enterprises. Inspur independently developed China's first mainframe for key applications: TSK1 which makes China, after the United States and Japan, the third country mastering core technologies for high-end servers. TSK1 also won the National Science and Technology Progress Award in 2014.

As one of China's top IT companies, for over 70 years, Inspur has been committed to becoming an advanced IT product manufacturer and a leading IT solution supplier that leads information trend and boosts civilization.

贵州中科生态云大数据科技有限公司

贵州中科生态云大数据有限公司是一家专业经营高科技产品企业。其中部门中经营分析检测、植物化学、化学合成及分子生物学等实验室通用进口仪器和耗材的公司。

公司以服务贵州科研院所和高等院校等用户为目标，销售代理国际先进的仪器和耗材系统，使科学研究、分析检测的方法及过程更为简便、可靠和快捷，旨在通过新方法和新产品来满足实验室全球化技术发展及国内日益需求的食品安全、医药卫生事业。

主要经营产品

美国 --- 珀金埃尔默 PE　　贵州市场代理
美国 --- 宜曼达测序仪　　贵州产品销售
美国 --- 安捷伦生物　　贵州产品推广商
美国 --- 美国生命技术有限公司　贵州产品销售

LAB-BIOGEN （混匀仪；掌上离心机；颗粒圆整度测定仪）
Biomerieux （全自动免疫荧光分析仪；空气采样器；全自动重量稀释仪；拍打式均质仪）
海尔 (超低温冰箱；生物安全柜）
Roche（荧光定量 PCR 仪；核酸分离纯化系统；测序仪）
Merck-Millipore（实验室纯水；超纯水；中央供水；过滤；超滤；完整性检测；水质检测；微生物检测）
Esco（生物安全柜；洁净工作台；CO2 箱；超低温冰箱）
Eppendorf（移液器；分配器；助吸器；吸头；分液管；离心管；冰盒；细胞培养耗材）

本公司在经营策略、方法、产品和解决方案中坚持以客户需求为中心原则、积极把新技术、新应用推广到需求用户手上。
让新科研新技术及相关科技术，促进用户的成功。

贵州合作伙伴：中科院贵州地球化学所、贵州医科大学、
遵义医学院、贵州大学、贵州食品药品监督管理局、
贵州省产品质量检测中心、农科院等

贵州师范大学生命科学学院简介

贵州师范大学生命科学学院前身为贵阳师范学院生物系，是全国较早的生物学教师教育专门机构之一。

1956 年由著名鱼类学家、时任贵阳师范学院副院长顾光中教授提议创办生物科，1958 年 8 月生物科转为生物系，顾光中教授任首任系主任。1994 年生物系更名为生物科学技术系。2003 年 7 月，学校进行教学科研组织机构调整，在生物技术专业基础上组建生物技术与工程学院，生物科学专业调整到地理与生物科学学院。2008 年 8 月，学校再次进行组织机构调整，撤消生物技术与工程学院、地理与生物科学学院，整合生物学相关专业，组建生命科学学院。

学院现设有生物科学、生物技术、生物工程和园艺 4 个本科专业（2009 年起至今实行大类招生，并在大二根据学生兴趣自主选择，专业分流）；拥有生物学和生态学两个一级学科硕士学位授权点和生物教育硕士专业学位授权点；拥有生物学和生态学两个贵州省特色重点学科；建设了贵州省植物生理与发育调控重点实验室、西南喀斯特山地生物多样性保护国家林业局重点实验室、国家生物学实验教学示范中心；打造了教育部喀斯特山地生物多样性保护与可持续利用创新人才团队，贵州省特色植物资源保护与可持续利用科技创新人才团队，生物科学专业基础课程省级教学团队和生物技术核心课程教学团队校级教学团队。生物科学专业 2007 年获批国家特色专业。生物技术专业 2008 年获批贵州省示范专业，2012 年获批国家级专业综合改革试点项目（第二批）。

学院在师资队伍建设方面，培养与引进并举，近年来从美国普渡大学、厦门大学、四川大学、重庆大学、西南大学、华中农业大学、山东农业大学、中国科学院等国内外知名高校和科研院所引进了一批优秀人才。学院现有教职工 83 人，其中专任教师 67 人，博士生导师 1 人、硕士生导师 30 人，教授 15 人，教师中有教育部高等学校生物科学与工程教学指导委员会委员 1 人，贵州省教学名师 2 人；多人先后获得“全国师德先进个人”“新世纪百千万人才工程奖”“省五四青年奖章”“省青年科技奖”“省优秀青年科技工作者”“省优秀教师”和“省优秀党务工作者”等光荣称号。目前，学院的教学科研人员中具有博士学位的占 80.6%，行政管理与教辅人员中具有硕士以上学位和

研究生学历的占87.5%；已经形成一支学历层次高、发展潜力大、年龄结构合理的人才队伍。

生命科学学院依托省级特色重点学科（生物学、生态学）和一级学科硕士授权点（生物学、生态学），搭建了由重点实验室、专业实验室和基础实验室组成的三级校内实践教学平台，完善了多层次人才培养条件。现有2个省部级重点实验室（贵州省植物生理与发育调控重点实验室、西南喀斯特山地生物多样性保护国家林业局重点实验室）、1个省级专业实验室（生物工程专业实验室）和2个省级基础实验室（生物技术基础实验室、基础生物学实验室）。拥有国家级生物学实验教学示范中心1个。实验教学用房约15000平方米，实验室仪器设备共825台件，固定资产总值逾1500多万元。此外，通过多年建设，结合贵州省生物技术产业的实际，与省内教育、科研机构、事业单位和企业共同建立了二十余个校外实习、实训基地，涉及基础教育、医药、食品、环保、农、林、牧、渔等行业，满足学生多元化实践的需求。

经过多年来的建设和发展，学院现已形成了喀斯特生物多样性与自然保护、喀斯特植物生理生态、等特色研究方向，并以科研反哺教学。近五年来，学院教师共主持承担了973、国家自然科学基金、国家林业局、国家环保部、贵州省省长基金、贵州省科技厅、贵阳市科技局等各级各类项目，近5年获国家级项目33项、省部级项目95项、厅级项目20项、横向项目36项；出版著作、教材等14部；授权国家发明专利13项。发表学术论文近五百篇，其中SCI、EI收录五十余篇。获贵州省科技进步奖3项（其中二等奖1项，三等奖2项），获贵州省高等学校教学成果奖2项（一等奖1项，二等奖1项）；承办了“生物医学大数据国际研讨会”“生态文明贵阳国际论坛2018年年会绿色发展与生物多样性保护研讨会”等学术交流活动。

贵州师范大学生命科学学院倡导教学、科研并重，实施开放式人才培养模式，因材施教，注重培养学生良好的国际化视野、独立的创新精神和较强的专业实践能力，确保毕业生就业途径全方位、多渠道,为贵州省教育事业和地方经济社会发展做出了巨大的贡献。未来，学院将进一步开拓创新、深化改革，进一步规划和凝练科研方向，不断巩固和彰显人才培养和科学研究的特色，继续深化应用型高级创新人才培养体系建设，加强内涵发展，不断提高教学科研水平和人才培养质量。面向贵州，积极主动为贵州社会

经济发展服务，朝着建设有地方教师教育特色多学科发展的学院的目标不断前进。