SCIENCE CHINA
Life Sciences

THEMATIC ISSUE: Computational life sciences
• EDITORIAL •

November 2014 Vol.57 No.11: 1051–1053
doi: 10.1007/s11427-014-4755-6

Systems biology and metagenomics: a showcase of Chinese bioinformatics researchers and their work

ZHU DongXiao[1*] & QIN Zhaohui S.[2*]

[1]*Department of Computer Science, Wayne State University, Detroit, MI 48084, USA;*
[2]*Department of Bioinformatics and Biostatistics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA*

Received September 10, 2014; published online October 14, 2014

Citation: Zhu DX, Qin ZS. Systems biology and metagenomics: a showcase of Chinese bioinformatics researchers and their work. Sci China Life Sci, 2014, 57: 1051−1053, doi: 10.1007/s11427-014-4755-6

It is our great honor to guest-edit this Thematic Issue on Bioinformatics in *Science China Life Sciences*. In recent years, a strong cohort of Chinese scientists has emerged as leading scholars in the exciting fields of bioinformatics and computational biology. In this issue, we are pleased to present outstanding research work produced from 10 world renowned researchers.

System biology and metagenomics are recently revitalized by the inception of next-generation sequencing technologies. The former seeks to understand the function and behavior of complex biology systems through studying the interaction of their components. The latter extends the systems biology concept and employs systems biology approaches to study the organization and interaction among microbes in their original habituate. In this Thematic Issue, we present a showcase of research work conducted by prominent researchers which can be broadly categorized into these two following areas.

(i) Systems biology to identify disease genes or biomarkers. Chen et al. [1] used graphical model approach to identify disease genes from biological networks. In their approach, they proposed a weighted kernel based Markov random field method and employed three different kernels to describe the overall relationship among the network nodes integrated from five networks. The disease genes were prioritized and selected using the posterior probability obtained from an improved Gibbs sampling procedure. The disease gene identification using topological features represents a promising direction. Given the increasing amount of functional annotation information available, Li et al. [2] developed a new approach to identify disease genes using both topological features and functional similarity measured by gene ontology (GO). In their approach, they used shortest-path algorithm to prioritize the disease gene in the protein-protein interaction network by integrating the semantic similarity of GO annotations. The idea of incorporating functional annotations is further implemented in a genome-wide association study (GWAS) to identify the disease associated single nucleotide polymorphisms (SNPs). As one of the representative studies, Lin et al. [3] proposed a statistical framework to estimate the extent of the improvement in disease SNP identification by incorporating functional annotation data into GWAS study. Although the improvement is not as significant as expected, it warrants further investigations in this promising direction.

The aforementioned methods can be considered as computational approaches that are applicable to the data collected from steady-state disease or homogeneous disease. It is well known that many diseases such as cancer are staged and/or stratified. Computational methods for identification of biomarkers from these diseases are in great demand. Zhang et al. [4] proposed a generative model to identify disease genes by analyzing topological variety among many protein-protein interaction networks. In their approach, they first learned common representation of multiple networks, i.e., hidden features, using a systematic feature memory

*Corresponding author (email: dzhu@wayne.edu; zhaohui.qin@emory.edu)

framework, then followed by a rebuilt of original network. The generative model approach has been shown to be effective using validation studies. Yu et al. [5] conducted a more comprehensive study that systematically compared algorithms for identifying disease biomarkers from dynamic networks. They found the top performed methods shared a substantial fraction of genes but not other methods. This study points out a challenge that how to integrate and use the results from different methods. A promising attempt in response to this challenge is instead of using single gene as biomarkers, network edges connecting a pair of genes can also be used as biomarkers. The recent studies in Zeng et al. [6] have demonstrated that edge biomarkers are more stable and consistent across the studies.

(ii) Metagenomics to identify microbe interaction, mechanism and outcomes. The declining cost of next-generation sequencing data and advances in systems biology have greatly boosted the field of metagenomics. Hundreds of millions of short reads can be sequenced from uncultivated bacterial samples from their natural habitat. Using these data, it is now possible to better answer a series of classical questions in metagenomics that have not been answered before, such as who they are and what they do. Jiang et al. [7] proposed a computational approach to investigate the microbial interaction networks existing among the microbes. The microbial interactions are critical to understand a series of outcomes of interest, for example, disease outcome. Ma et al. [8] looked at the problem from a fresh angle at biological pathway level; they attempted to explain the observed commonalities and differences in the genomic organizations of genes encoding specific pathways across different genomes. They reported two key observations: one is the frequencies of the transcription activation of pathways relative to those of the other encoded pathways in an organism; another is the variation in the activation frequencies of a specific pathway across the related genomes. It seems clear that molecular mechanisms including transcription and translation are central in studying the co-habitant of the microbes in their natural habitat. Yu et al. [9] recently conducted an in-depth study of molecular mechanism from a structural biology perspective, whose impact is expected to go far beyond metagenomics. Many computational approaches to study metagenomics are supervised methods meaning that they are capable of identifying known species and their interaction and organization. For large majority of unknown species, sequence assembly is a viable direction to pursue in metagenomics. In Chin et al. [10], they gave a comprehensive survey of the recent advances in sequencing assembly algorithms and pointed out challenges and potential solutions.

This Thematic Issue has collected some of representative research advance in systems biology and metagenomics. Among many potential directions, synergistic area such as metagenomic systems biology that uses systems biology approaches to study metagenomics is among the promising directions to explore.

1 Chen BL, Li M, Wang JX, Wu FX. Disease gene identification by using graph kernels and Markov random fields. Sci China Life Sci, 2014, 57: 1054–1063
2 Li M, Li Q, Ganegoda GU, Wang JX, Wu FX, Pan Y. Prioritization of orphan disease-causing genes using topological feature and GO similarity between proteins in interaction networks. Sci China Life Sci, 2014, 57: 1064–1071
3 Hou L, Ma TZ, Zhao HY. Incorporating functional annotation information in prioritizing disease associated SNPs from genome wide association studies. Sci China Life Sci, 2014, 57: 1072–1079
4 Zhang Y, Cheng Y, Jia KB, Zhang AD. A generative model of identifying informative proteins from dynamic PPI networks. Sci China Life Sci, 2014, 57: 1080–1089
5 Yu H, Mitra R, Yang J, Li Y, Zhao ZM. Algorithms for network-based identification of differential regulators from transcriptome data: a systematic evaluation. Sci China Life Sci, 2014, 57: 1090–1102
6 Zeng T, Zhang WW, Yu XT, Liu XP, Li MY, Liu R, Chen LN. Edge biomarkers for classification and prediction of phenotypes. Sci China Life Sci, 2014, 57: 1103–1114
7 Jiang XP, Hu XH. Inferring microbial interaction networks based on consensus similarity network fusion. Sci China Life Sci, 2014, 57: 1115–1120
8 Ma Q, Chen X, Liu C, Mao XZ, Zhang HY, Ji F, Wu CG, Xu Y. Understanding the commonalities and differences in genomic organizations across closely related bacteria from an energy perspective. Sci China Life Sci, 2014, 57: 1121–1130
9 Yu DM, Zhang C, Qin PW, Cornish VP, Xu D. RNA-protein distance patterns in ribosomes reveal the mechanism of translational attenuation. Sci China Life Sci, 2014, 57: 1131–1139
10 Chin FYL, Leung HCM, Yiu SM. Sequence assembly using next generation sequencing data—challenges and solutions. Sci China Life Sci, 2014, 57: 1140–1148

Biographical Sketch

Zhu DongXiao is currently an Assistant Professor at Department of Computer Science, Wayne State University. From 2008 to 2011, he was an Assistant Professor at Department of Computer Science, University of New Orleans. From 2006 to 2008, he worked at Stowers Institute for Medical Research as a Biostatistician. He received his Ph.D. from University of Michigan in 2006. His research interests have been in areas of computational biology, bioinformatics, health informatics and the interface with data mining, machine learning and pattern recognition. Dr. Zhu has published nearly 40 peer-reviewed publications and numerous book chapters and he served on several editorial boards of bioinformatics journals. Dr. Zhu's research has been supported by National Institutes of Health (NIH), National Science Foundation (NSF), State of Louisiana and private agencies and he has served on multiple NIH and NSF grant review panels. Dr. Zhu has advised numerous students at undergraduate, graduate and postdoctoral levels.

Biographical Sketch

Qin Zhaohui S. is currently an Associate Professor in the Department of Biostatistics and Bioinformatics at Rollins School of Public Health, Emory University. He is also a faculty member at the Department of Biomedical Informatics, Emory University School of Medicine and Biostatistics and Bioinformatics Shared Resource, Winship Cancer Institute. Dr. Qin received his B.S. degree in Probability and Statistics from Peking University in 1994 and Ph.D. degree in Statistics from University of Michigan in 2000. He was a postdoctoral fellow in Dr. Liu Jun's group in Department of Statistics at Harvard University from 2000 to 2003. He joined the Department of Biostatistics at University of Michigan in 2003. In 2010, he moved to his current position in Emory University. Dr. Qin has more than ten years of experience in statistical modeling and statistical computing with applications in statistical genetics and genomics. Recently, his research is focused on developing Bayesian model-based methods to analyze data generated from applications of next-generation sequencing technologies such as ChIP-seq, RNA-seq and resequencing. Dr. Qin also actively collaborates with biomedical scientists and clinicians on various projects that utilize next-generation sequencing technologies to study cancer genomics. Dr. Qin has published more than 70 peer-reviewed research papers covering statistics, bioinformatics, statistical genetics and computational biology. He has supervised more than 10 graduate students and postdoctoral fellows.