

## Editorial

# Functional Genomics, Genetics, and Bioinformatics 2016

**Youping Deng,<sup>1</sup> Hongwei Wang,<sup>2</sup> Ryuji Hamamoto,<sup>3</sup> Shiwei Duan,<sup>4</sup>  
Mehdi Pirooznia,<sup>5</sup> and Yongsheng Bai<sup>6</sup>**

<sup>1</sup>Bioinformatics Core, Office of Biostatistics & Quantitative Health Sciences, Department of Tropical Medicine, Medical Microbiology, and Pharmacology, University of Hawaii John A. Burns School of Medicine, Honolulu, HI 96813, USA

<sup>2</sup>Department of Medicine, University of Chicago, Chicago, IL 60637, USA

<sup>3</sup>Division of Molecular Modification and Cancer Biology, National Cancer Center Research Institute, Tokyo, Japan

<sup>4</sup>School of Medicine, Ningbo University, Ningbo, Zhejiang 315211, China

<sup>5</sup>Bioinformatics and Computational Biology Core Facility, National Heart, Lung, and Blood Institute, Office of the Scientific Director, National Institutes of Health, Bethesda, MD 20814, USA

<sup>6</sup>Department of Biology, Indiana State University, Terre Haute, IN 47809, USA

Correspondence should be addressed to Youping Deng; [dengy@hawaii.edu](mailto:dengy@hawaii.edu)

Received 17 October 2016; Accepted 17 October 2016

Copyright © 2016 Youping Deng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

During the postgenomics era, more and more “omics” data are being produced due to rapid development of cutting-edge technologies such as next generation sequencing. These omics data include genomics [1–3], transcriptomics [4–7], proteomics [8–10], metabolomics [11–13], and epigenomics [14–16]. Bioinformatics plays a central role in analyzing and managing this huge amount of “omics” data, further understanding the function of biological molecules in different levels.

Several bioinformatics tools and methods have been developed in the special issue. C.-H. Tung et al. developed a new method called QuaBingo, a prediction system for protein quaternary structure attributes using block composition. The method is 23% of Matthews Correlation Coefficient (MCC) higher than the existing prediction systems. *Exiguobacterium antarcticum* B7 is extremophile Gram-positive bacteria able to survive in cold environments. A key factor to understanding cold adaptation processes is related to the modification of fatty acids composing the cell membranes of psychrotrophic bacteria. R. Kawasaki et al. reconstructed the fatty acid biosynthesis pathway of *E. antarcticum* B7 based on both genomic and bibliomic data using bioinformatics methods, which is a great resource for the research of *Exiguobacterium antarcticum* B7. L. Hua and C. Quan built a novel method for protein-protein interaction (PPI) extraction using a shortest dependency path based CNN (sdpCNN)

model. The proposed method only takes the sdp and word embedding as input and could avoid bias from feature selection by using CNN. The new approach outperformed traditional state-of-the-art kernel based methods. Clustered regularly interspaced short palindromic repeat (CRISPR) is a genetic element with active regulation roles for foreign invasive genes in the prokaryotic genomes and has been engineered to work with the CRISPR-associated sequence (Cas) gene Cas9 as one of the modern genome editing technologies. G. Mai et al. provided a valuable comprehensive curation resource to show the dynamic evolutionary patterns of prokaryotic CRISPRs based on computational evolutionary analysis of 8 completely sequenced species in the genus *Thermoanaerobacter*.

Two papers are focused on transcriptomics data analyses. Q. Sun et al. tried to understand the gene function of thgal in *Trichoderma harzianum* Th-33, important biocontrol filamentous fungi, which are widely used for their adaptability, broad antimicrobial spectrum, and various antagonistic mechanisms. Illumina RNA-seq technology (RNA-seq) was used to determine transcriptomic differences between the wild-type strain and thgal mutant. A total of 888 genes were identified as differentially expressed genes (DEGs), including 427 upregulated and 461 downregulated genes. According to the functional annotation of these DEGs, they found the most abundant group was “secondary metabolite

biosynthesis, transport, and catabolism.” Hepatitis E virus-(HEV-) mediated hepatitis has become a global public health problem. K. Xu et al. investigated the function of ORF3 from the swine form of HEV (SHEV); high-throughput RNA-Seq-based screening was conducted to identify the differentially expressed genes in ORF3-expressing HepG2 cells. The results indicated that, in the established ORF3-expressing HepG2 cells, the mRNA levels of CLDN6, YLPM1, APOC3, NLRP1, SCARA3, FGA, FGG, FGB, and FREM1 were upregulated, whereas the mRNA levels of SLC2A3, DKK1, BPIFB2, and PTGRI were downregulated.

Several studies are concentrated on functional genomics data analyses. Using Web service SNP\_TATA\_Comparator presented in their previous paper, P. Ponomarenko et al. analyzed immediate surroundings of known SNP markers of diseases and identified 53 candidate SNP markers that can significantly change the affinity of TATA-binding protein for human gene promoters, with circadian consequences. These candidate SNP markers could be potentially useful for physicians (to select optimal treatment for each patient) and for the general population (to choose a lifestyle preventing possible circadian complications of diseases). M. Abu Saleh et al. conducted a comprehensive computational analysis on the functional and structural impacts of single nucleotide polymorphisms (SNPs) of the human ADIPOR1 at protein level. Their analyses suggested that the aforementioned variants, especially H341Y, could directly or indirectly destabilize the amino acid interactions and hydrogen bonding networks of ADIPOR1. Z.-X. Chen et al. used BLAST to call SNPs for non-model organisms based on 16 mixed functional gene's sequence data of polyploidy wheat. They demonstrated that mixed samples' NGS sequences and then analysis by BLAST were an effective, low-cost, and accurate way to mine SNPs for nonmodel species. Assembled reads and polynomial fitting threshold were recommended for more accurate SNPs targets.

One article deals with proteomics data analysis. S. Wan et al. identified important differential proteins between patients of slow transit constipation and normal controls using two-dimensional electrophoresis followed by laser desorption ionization tandem time-of-flight mass spectrometry (MALDI-TOF-MS). One paper is related to epigenomics data analysis. MiR-23a-27a-24-2 cluster has various functions and aberrant expression of the cluster is a common event in many cancers. Y. Wang et al. found a CG-rich region spanning two SP1 sites in the cluster promoter region. The SP1 sites in the cluster were demethylated and methylated in Hep2 cells and HEK293 cells, respectively. The demethylated SP1 sites in miR-23a-27a-24-2 cluster upregulate the cluster expression, leading to proliferation promotion and early apoptosis inhibition in laryngeal cancer cells.

Interaction of gene and environmental factors plays an important role in human diseases. Y. Wang et al. have found many important risk factors that affect chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS), including biological, social, and psychological factors. They also discussed the potential interaction between genes and these risk factors.

In summary, this special issue presents a broad range of topics from functional genomics, transcriptomics, proteomics, epigenomics, and bioinformatics. It covers a variety of diseases such as cancer, hepatitis, chronic prostatitis, and infectious diseases. The study organisms include human, plant, and microorganisms. We hope that the readers will find interesting knowledge and methods in the issue.

## Acknowledgments

The work edition was supported by the Grant NIH 5P30GM114737 and the Grant NIH P20GM103466.

Youping Deng  
Hongwei Wang  
Ryuji Hamamoto  
Shiwei Duan  
Mehdi Pirooznia  
Yongsheng Bai

## References

- [1] M. Jia, Y. Liu, Z. Shen et al., “HDAM: a resource of human disease associated mutations from next generation sequencing studies,” *BMC Medical Genomics*, vol. 6, supplement 1, p. S16, 2013.
- [2] B. Bonev and G. Cavalli, “Organization and function of the 3D genome,” *Nature Reviews Genetics*, vol. 17, no. 11, pp. 661–678, 2016.
- [3] X. Chen, Y. Xu, W. Yang et al., “Association of six CpG-SNPs in the inflammation-related genes with coronary heart disease,” *Human Genomics*, vol. 10, supplement 2, p. 21, 2016.
- [4] L. Hu, J. Ai, H. Long et al., “Integrative microRNA and gene profiling data analysis reveals novel biomarkers and mechanisms for lung cancer,” *Oncotarget*, vol. 7, no. 8, pp. 8441–8454, 2016.
- [5] M. Uhlén, B. M. Hallström, C. Lindskog, A. Mardinoglu, F. Pontén, and J. Nielsen, “Transcriptomics resources of human tissues and organs,” *Molecular Systems Biology*, vol. 12, no. 4, p. 862, 2016.
- [6] X. Wei, J. Ai, Y. Deng et al., “Identification of biomarkers that distinguish chemical contaminants based on gene expression profiles,” *BMC Genomics*, vol. 15, article 248, 2014.
- [7] Y. Deng, S. A. Meyer, X. Guan et al., “Analysis of common and specific mechanisms of liver function affected by nitrotoluene compounds,” *PLoS ONE*, vol. 6, no. 2, Article ID e14662, 2011.
- [8] F. Zhang, Y. Deng, M. Wang, L. Cui, and R. Drabier, “Pathway-based biomarkers for breast cancer in proteomics,” *Cancer Informatics*, vol. 13, supplement 5, pp. 101–108, 2014.
- [9] D. Kumar, G. Bansal, A. Narang, T. Basak, T. Abbas, and D. Dash, “Integrating transcriptome and proteome profiling: strategies and applications,” *Proteomics*, vol. 16, no. 19, pp. 2533–2544, 2016.
- [10] J. Wang, M. Li, Y. Deng, and Y. Pan, “Recent advances in clustering methods for protein interaction networks,” *BMC Genomics*, vol. 11, supplement 3, p. S10, 2010.
- [11] X. Chen, H. Chen, M. Dai et al., “Plasma lipidomics profiling identified lipid biomarkers in distinguishing early-stage breast cancer from benign lesions,” *Oncotarget*, vol. 7, no. 24, pp. 36622–36631, 2016.

- [12] M. Marcinkiewicz-Siemion, M. Ciborowski, A. Kretowski, W. J. Musial, and K. A. Kaminski, "Metabolomics—a wide-open door to personalized treatment in chronic heartfailure?" *International Journal of Cardiology*, vol. 219, pp. 156–163, 2016.
- [13] A. Cambiaghi, M. Ferrario, and M. Masseroli, "Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration," *Briefings in Bioinformatics*, 2016.
- [14] J. Melson, Y. Li, E. Cassinotti et al., "Commonality and differences of methylation signatures in the plasma of patients with pancreatic cancer and colorectal cancer," *International Journal of Cancer*, vol. 134, no. 11, pp. 2656–2662, 2014.
- [15] Z. Sun, J. Cunningham, S. Slager, and J. P. Kocher, "Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis," *Epigenomics*, vol. 7, no. 5, pp. 813–828, 2015.
- [16] J. Zierer, C. Menni, G. Kastenmüller, and T. D. Spector, "Integration of 'omics' data in aging research: from biomarkers to systems biology," *Aging Cell*, vol. 14, no. 6, pp. 933–944, 2015.