

Editorial

Special collection of recent advances in next-generation bioinformatics

Xuegong Zhang*

Ministry of Education Key Laboratory of Bioinformatics and Bioinformatics Division, Center for Synthetic and Systems Biology, Tsinghua National Laboratory for Information Science and Technology/Department of Automation, Tsinghua University, Beijing 100084, China

* Correspondence: zhangxg@tsinghua.edu.cn

Received February 6, 2016

The National Basic Research Program of China, also known as the 973 Program, has been the most prestigious center grant for basic research of various fields funded by the central government in China for the past 20 years. In this program, we have been lucky to have the opportunity to conduct a 5-year research project (#2012CB316500) on “Bioinformatics Theories and Methods for Next-Generation Sequencing” since 2012. We call it the project on Next-Generation Bioinformatics for short. The project targeted at several major challenges in two aspects of the analysis of big data in biological and medical research introduced by the rapid development and application of next-generation sequencing. One is the methodology and computing power in handling the big sequencing data and other omics data, and the other is the detection of patterns, discovery of knowledge and building of models based on the complex data.

The project is organized as five subprojects undertaken by 30 co-PIs from 9 institutions. I have been the PD (Program Director) for the whole project as well as the PI for subproject 4. The five subprojects are: (i) Data generation models of next-generation sequencing and optimization of data processing methods, led by Dr. Xuan Li of Shanghai Institute of Biological Sciences, Chinese Academy of Sciences; (ii) Computational models and architectures for next-generation sequencing big data, led by Dr. Peiheng Zhang of the Institute for Computing Technology, Chinese Academy of Science; (iii) Next-generation sequencing-based epigenomics study for

cell differentiation and cancer development, led by Dr. Michael Q. Zhang of Tsinghua University; (iv) Theory and methods for next-generation transcriptome data and integrative network analysis, led by Dr. Xuegong Zhang of Tsinghua University; and (v) Statistical genetics theory and methods based on next-generation sequencing data, led by Dr. Zewei Luo of Fudan University. Co-PIs of the whole team come from Tsinghua University, Fudan University, Southeast University, The Second Military Medical University, Shanghai Institute of Biological Sciences (SIBS) of CAS, Institute of Computing Technology (ICT) of CAS, Beijing Institute of Genomics of CAS, CAS-MPG Partner Institute for Computational Biology (PICB), and Shanghai Center for Bioinformation Technology (SCBT).

The past few years have evidenced rapid development in the field of quantitative biology especially in areas related with genomics and next generation sequencing. Our team has achieved a series of scientific, technical and methodology advances in all aspects of the project in the past four years. The achievement covers methods and algorithms for next-generation sequencing data, epigenomics especially about 3D structure of genomes and long-range chromatin interactions, gene expression and splicing analysis, understanding on the function and regulation of long non-coding RNAs, methods for analyzing biological pathways and networks, mathematical models of sequencing data and computing architecture for massive genomic data, methods in sequencing-based genetics and evolu-

tion analysis, quantitative modeling of gene regulation, methods for omics data integration and application on cancer study, etc.

As a part of the conclusion of this project, we invited co-PIs to submit reviews on recent progresses in their fields as a special collection of *Quantitative Biology* dedicated to this project. The submissions were peer reviewed in the same procedure as standard submissions to *Quantitative Biology*. Starting from this issue, the accepted submissions from this project after the peer review will be published as papers in this special collection. Four papers in the collection are published in this issue. The paper “*Mapping and differential expression analysis from short-read RNA-seq data in model organisms*” by Zhao *et al.* from subproject 1 reviewed recent developments in methods and algorithms for RNA-seq data processing, for the scenario of model organisms where reference genomes can be used. In “*Combinatorial pooled sequencing: experiment design and decoding*”, Cao and Sun from subproject 1 reviewed the principle and methods for the strategy of sequencing multiple samples by sequencing pooled samples of some special “smart” combination design instead of sequencing each single sample. The strategy provides an interesting and promising example of using information theory to guide the efficient design of

biological experiments. The rapid development of next-generation sequencing is enabling more and more accurate and detailed measurements of transcriptomes, which make the quantitative modeling of complex molecular pathways and systems possible. In “*A quantitative understanding of microRNA-mediated competing endogenous RNA regulation*”, Yuan *et al.* from subproject 3 reviewed the recent development of mathematical models that capture the law of microRNA regulation. This is a good example of how high-throughput experimental technology and bioinformatics analysis can advance our understanding of complex biology processes from qualitative to quantitative, which is a key feature of the next-generation bioinformatics. Another trend in next-generation bioinformatics is the integrative study of multiple types of data, which is a big challenge both for information science and biological science. Wang and Gu from subproject 3 reviewed the progress in an important area of cancer biology in the paper “*Integrative clustering methods of multi-omics data for molecule-based cancer classification*”. We hope these and the future papers in the special collection will not only show the coverage and progress of our 973 project, but also bring to the community our understanding and perspective on the future development of next-generation bioinformatics.