EDITORIAL

Special collection of recent advances in next-generation bioinformatics, part II

Xuegong Zhang*

MOE Key Laboratory of Bioinformatics, Bioinformatics Division and Center for Synthetic & Systems Biology,TNLIST; Department of Automation, Tsinghua University, Beijing 100084, China
* Correspondence: zhangxg@tsinghua.edu.cn

Recent advances in quantitative biology has attested the important roles bioinformatics play in biological sciences as well as the great expansion of the scope of next-generation bioinformatics to all areas of systems biology. From the last issue of Quantitative Biology, we started to publish a special collection of reviews and research articles on next-generation bioinformatics to reflect this trend. The collection is dedicated to the completion of the 5-year center grant (#2012CB316500) on next-generation bioinformatics under the National Basic Research Program of China or the "973 Program". The project aims to advance research on basic bioinformatics methodologies for processing next-generation biological data especially multiple types of sequencing data, and for converting the information buried in the data into quantitative understanding of important biological processes. The papers in the special collection were selected from submissions by the principle investigators and their collaborators of this project, after the standard peer-review procedure of Quantitative Biology. They reflect current progresses from the team and the whole community in several major directions that the project covers.

The special collection will be continued in this issue and the next issue. Four papers are published in this issue. In the last issue, Zhao *et al*. reviewed methods for processing RNA-sequencing data for transcriptomes for which the reference genome is available. In this issue, Li *et al*. from the subproject 1 told the other aspect of the story by reviewing methods for processing RNA-sequencing data when the reference genome is not available or complete. The title of the paper is "*De novo assembly of transcriptome from next-generation sequencing data*". In genomic sequencing analysis for genetics studies, obtaining haplotype data is an important and difficult task. In the paper "*Comparison of the experimental methods in haplotype sequencing via next generation sequencing*", Tu *et al*. from subproject 1 reviewed major technologies for haplotype sequencing, and compared their performances and characteristics. Wu *et al*. from subproject 4 reviewed major sequencing platforms and bioinformatics strategies and methods for medical genetics studies based on next-generation sequencing, with the title "*Whole genome sequencing and its applications in medical genetics*". It provided a rich source for bioinformatics tools available in this field and discussed existing open issues and possible future directions for dealing with the issues. Genomes are organized in three-dimensional space rather than one-dimensional linear space. Although analyzing the 1D genome sequence is fundamental in genetics studies, obtaining the 3D structure is crucial for understanding many key aspects of the genome organization and its relation to higher order gene regulation. In the paper "*Developing bioimaging and quantitative methods to study 3D genome*", Gao *et al*. from subproject 3 reviewed existing knowledge on the structural organization of genomes, and the advances in super-resolution microscopy techniques for directly detecting 3D structure of a genome. Progresses on software for processing and visualizing Hi-C sequencing data are also introduced, followed by discussions on future trends for integrating image-based approaches and sequencing-based approaches for better understanding the 3D genome organization.

Another four papers of the special collection will be published in the next issue. In the paper "*An overview of*

major metagenomic studies on human microbiomes in health and disease", Cui *et al*. from subproject 4 provide a comprehensive overview of the major types of studies on human microbiomes based on recent sequencing and bioinformatics technologies. It provides a rather complete survey on what have been done on microbiomes in and on various sites of the human body for people in different health conditions, and also provides a rich collection of available databases and data archives in this field. Open questions and future research directions are also discussed. In the paper "*Design of efficient simplified genomic DNA and bisulfite sequencing in large plant populations*", Wu *et al*. from subproject 5 propose an optimized experiment design of simplified experiments for genomic DNA and bisulfite sequencing. It can greatly improve the efficiency of sequencing experiments, and is particularly useful for genotyping large plant populations. In another paper "*Revisiting the false positive rate in detecting recent positive selection*" from subproject 5, Xiang-Yu *et al*. address an important theoretical topic of the detection of positive selection events from genetic data. It discusses sources of possible false positives in the detection, and strategies for developing more reliable tests. In the paper "*Advances in computational ChIA-PET data analysis*", He *et al*. from subproject 3 compare the characteristics of ChIA-PET sequencing for analyzing 3D genome structure with the Hi-C sequencing method, and review new methods developed from the team and others for better analyzing ChIA-PET data.

These will complete the special collection of 12 papers, covering a wide range of topics from algorithms and methods for processing multiple types of next-generation sequencing data, to the functional or evolutionary studies of important biological processes and their quantitative models. I would like to thank all the authors for their generous contributions to the special collection, and thank all the reviewers for their timely and insightful comments and suggestions on the submissions. The completion of the 5-year 973 project on next-generation bioinformatics will mark great progresses in the methodological researches and applications across all these fields, and also open the door to more exciting advances and discoveries in the future research.