Editorial

Yan Guo and Quanhu Sheng

Department of Cancer Biology, Vanderbilt University, 2220 Pierce Ave, 494 Preston Research Building, Nashville TN, 37212, USA Email: Yan.guo@vanderbilt.edu Email: quanhu.sheng@vanderbilt.edu

Biographical notes: Yan Guo is an Assistant Professor at the Department of Cancer Biology and a Full-Time Faculty Bioinformatician in the Vanderbilt Center for Quantitative Sciences. He has been serving as the Technical Director of Bioinformatics for Vanderbilt Technologies for Advanced Genomics Analysis and Research Design. Prior to receiving his PhD, he gained laboratory research experience in conducting in vitro and in vivo pre-clinical studies at Vanderbilt Microarray Shared Resources. After joining the Faculty of Vanderbilt School of Medicine, his research has been focused on development of bioinformatics methodology and analysis approaches for cancer genomic studies.

Quanhu Sheng is a Research Assistant Professor in the Department of Cancer Biology and a Full-Time Faculty Bioinformatician in the Vanderbilt Center for Quantitative Sciences. He received his PhD in Bioinformatics at the Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences. He has expertise in next-generation sequencing data analysis, network and pathway analysis, proteomics/glycomics data analysis, and biological database construction. Until now, he has published 56 peer-reviewed papers, including 10 next generation sequencing related papers published in the last two years.

Computational biology and genomics refers to the use of computational analysis to decipher biology from genome sequences and related data, including both DNA and RNA sequences, as well as other 'post-genomic' data. The biomedical research field has progressed significantly over the last few years primarily due to the advancement of technologies such as next generation sequencing (NGS) and shotgun proteomics. Transforming genomic and proteomic information into biological knowledge requires creative and innovative new computational approaches. With the current abundance of massive biological datasets, computational studies have become one of the most important avenues for biological discovery.

After extensive peer review, we have selected seven manuscripts in computational genomics field for this special issue *Advances in Computational Biology and Genomics*. Here are the summary contributions for each of the manuscripts.

Wang et al. provided an in depth review of the latest proteomic technology: sequential window acquisition of all theoretical fragment ion spectra coupled to tandem

190 Y. Guo and Q. Sheng

mass spectrometry (SWATH-MS). SWATH-MS is a strategy for protein identification and quantification which operates in unbiased data-independent acquisition (DIA) mode and combines the advantages of two widely used MS-based techniques: shotgun (high throughput) and SRM proteomics (high sensitivity and reproducibility). This review illustrated the principle and implementation of SWATH-MS, and summarised its application as well as a perspective on the future of this powerful technology.

Sheng et al. introduced a high throughput sequencing data analysis pipeline NGSPERL. With the huge amounts of next generation sequencing data being generated, data analysis has become the bottle-neck of the research procedure. NGSPERL is a semi-automated module-based framework, which was especially designed for cluster-based high-throughput sequencing data analysis. User can either use recommended pipelines for RNA sequencing, exome sequencing, small RNA sequencing data, or build their own pipelines by plug multiple modules together. Such a flexible framework will significantly automate and simplify the process of large scale sequencing data analysis.

Velayutham proposed a rough neural network (RNN) classification method for mental imagery, multi-class electroencephalography dataset. The authors demonstrated improvement in term of classifying accuracy compared to back-propagation neural network algorithm.

Wajid et al. presented a modular approach to reference assisted genome assembly pipeline (MARAGAP) to determine the optimal reference sequence for assembly. The optimal reference sequence is used as a template to infer inversions, insertions, deletions and SNPs in the target genome. MARAGAP uses an algorithmic approach to detect and correct inversions and deletions, a De-Bruijn graph based approach to infer the insertions, an affinematch affine-gap local alignment tool to estimate the locations of insertions and a Bayesian estimation framework for detecting SNPs.

The study of Boucheham et al. is focused on a unique approach of biomarker discovery. In their approach, biomarker discovery is handled by means of a cooperative parallel and distributed approach based on metaheuristics. More specifically, metaheuristics are employed according to generalised island model architecture and a hybrid wrapper/filter mechanism is applied within the selection process. In addition, a new initialisation strategy based on an ensemble of filters is proposed to bootstrap the metaheuristics. The authors also used cancer datasets to demonstrate the improved performance.

Li and Scheich introduced a novel method for predicting lung cancer stages using a combination of differential expression analysis and machine learning techniques. And the authors show that their methods can predict lung cancer stage with 67–71% accuracy.

The study of Kaur et al. is related to the application of supervised learning techniques to generate a target-based predictive model for the knowledge base to foretell future advance in drug design. The objective of target-based modelling is prediction of the activity and relationships among different compounds from a large database with unknown activity and thus reducing the cost and time for discovery and development of a new drug.

The goal of this special issue is to stimulate the continuing effort in computational genomics for more efficient analysis on Omics data. We hope you enjoy the work performed by the researchers who contributed to this special issue.

Editorial

Acknowledgements

We would like to thank the anonymous reviewers for their careful review and critical comments which significantly improved the quality of the manuscripts in this special issue. Yan Guo and Quanhu Sheng were supported by Cancer Center Support Grant (P30CA068485).