Preface

Human Heredity

Hum Hered 2017/2018;83:105–106
DOI: 10.1159/000494818

Published online: January 9, 2019

Next-Generation Sequencing in Human Genetic Studies: Genome Technologies and Applications to Human Genetic Studies

Junwen Wang[a, b] Kai Wang[c, d] Xiaoming Liu[e] Pak Sham[f]
Zhongming Zhao[g, h]

[a]Department of Health Sciences Research, Center for Individualized Medicine, Mayo Clinic, Scottsdale, AZ, USA; [b]College of Health Solutions, Arizona State University, Phoenix, AZ, USA; [c]Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; [d]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA; [e]College of Public Health, University of South Florida, Tampa, FL, USA; [f]Centre for Genomic Sciences, Department of Psychiatry, State Key Laboratory in Cognitive and Brain Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong, PR China; [g]Department of Epidemiology, Human Genetics & Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA; [h]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA

This year marks the 70th anniversary of *Human Heredity*. The journal was first published in 1948, under the original title of *Acta Genetica*. Although the journal's name has changed, its mission statement of contributing to the advances in human genetics has remained the same. However, the field of human genetics today is much different from when the journal was founded, which was before the structure of DNA was discovered. The journal has seen all the major milestones in the field of human genetics over the past several decades, from the mapping of monogenic disease genes using linkage analysis, to the study of complex diseases – at first using candidate gene association analysis and more recently by more powerful and high-resolution genome-wide approaches. At the same time, our ability to evaluate the likely functional consequences of genetic changes has increased dramatically through the accumulation of genomic data on large population samples, as well as detailed functional genomic annotation on regulatory as well as coding sequences.

With these advances, we now have unprecedented opportunities to uncover the complex genetic underpinnings of a disease, paving the way to precision medicine. However, these advances also pose new challenges: how to process, analyze, and integrate large-scale, multi-dimensional, genomic data with heterogeneous, complex phenotypes; and how to deal with data missingness and false discovery of genetic signals. Thus, bioinformatics has become indispensable for human genetics research and has indeed developed as a field in its own discipline. In recognition of this, *Human Heredity* is publishing a special issue dedicated to bioinformatics, in particular covering the methods and applications that are relevant to human genetics. This marks the expansion of the focus of the journal from statistical and population genetics to

Junwen Wang
Department of Health Sciences Research, Center for Individualized Medicine
Mayo Clinic
13400 E Shea Blvd.
Scottsdale, AZ 85259 (USA)
E-Mail Wang.Junwen@Mayo.edu

Zhongming Zhao
School of Biomedical Informatics
The University of Texas Health Science Center at Houston
7000 Fannin St., Suite 820
Houston, TX 77030 (USA)
E-Mail Zhongming.Zhao@uth.tmc.edu

E-Mail karger@karger.com
www.karger.com/hhe

bioinformatics, genomics, and computational biology. It is our vision that such technologies and applications will advance *Human Heredity* research substantially.

In this special issue, we have 5 articles contributed from the International Conference on Intelligent Biology and Medicine (ICIBM 2018), held on June 10–12, 2018, in Los Angeles, CA, USA. The ICIBM received nearly 80 original research manuscripts. This issue covers not only traditional areas of genotype imputation, genetic and epigenetic interactions, but also cutting-edge technologies such as single-cell RNA sequencing, deep learning, and multi-omics data integration. These topics fit well with our newly expanded area of bioinformatics, genomics, and computational biology of the journal. We briefly introduce these 5 papers below.

Shi et al. systematically evaluated current tools for genotype imputation from genome-wide association (GWAS) data. They found most tools perform well, with IMPUTE2 performing slightly better, and recommended a minimum SNP density of 200 SNPs/Mb. More interestingly, they found that the majority of SNPs detected at 15× sequencing depths can be imputed from SNPs detected from 4× depths and the 1,000 genomes project. If this is true, we can significantly reduce our sequencing costs with more computing time. However, the authors tested this only on 10 Chinese individuals sequenced in the 1,000 genomes project, and the SNPs were called only by GATK. A thorough test on more populations, a larger sample size, and different SNP calling tools are needed to strengthen this conclusion.

Wang et al. developed a statistical framework to integrate data from GWASs, expression quantitative trait loci (eQTL), and protein-protein interactions (PPIs). Their method, called GeP-HMRF, is based on a Hidden Markov Random Field model, a well-known statistical model that can model PPIs. This method has some advantages when compared with the widely used method Sherlock. For example, it tests multiple genes in a PPI network, while Sherlock tests only one single gene each time. In the comparison of GeP-HMRF with Sherlock and COLOC methods using 9 GWAS datasets, the authors found GeP-HMRF could significantly improve the prediction accuracy. The software is deposited to open-source GitHub and is publicly accessible.

Kogan et al. applied a novel, nonparametric, gene-centric approach to test the interaction between SNPs and epigenetic CpG sites using the data from the Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE) project. They reported 12 genes with significant SNP-CpG interactions. Of these, three were previously implicated in asthma risk or underlying biological pathways. Their method integrates genetic (SNP) and epigenetic (CpG) data for disease gene prioritization, which helps the investigators to discover interactive genetic signals between genomic and epigenomic variations, leading to a better understanding of the pathophysiology of the disease.

Gao et al. introduced a deep learning approach for tRNA prediction. They built 13 models from 3 popular deep learning architectures: convolutional neural network (CNN), recurrent neural network (RNN), and a hybrid combination of both (CNN-RNN). Compared with existing state-of-the-art machine learning methods such as Support Vector Machine (SVM), K-nearest neighbors (KNN), and LibMutil ensemble classifier, their method performed the best among all evaluation metrics. Furthermore, their method can extract genomic features without extensive manual feature engineering, and their model substantially outperforms the widely used tRNAscan-SE method under various experiments. Deep learning methods are making headline news in imaging analysis and healthcare data mining, among others. As more and more large-scale genomic data are available, we urgently need breakthroughs of deep learning methods in sequencing data analysis. Deep learning will bring revolution to our traditional bioinformatics and genetics. It will dramatically accelerate our scientific discovery.

In the last paper, Huss et al. studied how exome capture kit affects the quality of single-cell whole exome sequencing (scWES). They used two kits, the Nextera rapid capture (NXT) from illumine Inc., which is recommended for scWES; and Agilent SureSelect XT Target Enrichment System (AGL), which is widely used for bulk sequencing. They found AGL outperformed NXT in coverage uniformity, mapping rates of reads, exome capture rates, and low PCR duplicate rates. Their data also showed that AGL outperforms NXT on both germline and somatic mutation detections. The results suggest that AGL may be used for scWES to produce better quality scWES data, though the kit is mainly designed for bulk samples.

These five papers are only a glimpse of the diverse and rapidly evolving field of bioinformatics and computational genomics. We welcome more submissions of high-quality manuscripts in areas such as precision medicine, methodology and algorithm development for next-generation sequencing data analysis (whole genome sequencing, whole exome sequencing, RNA-seq, ChIP-seq, methylation sequencing, etc.), multi-dimensional omics data integration, metagenomics, and other emerging areas such as immunotherapy, single-cell sequencing, and deep learning.