NUS National University of Singapore | NUS Centre for Cancer Research
Yong Loo Lin School of Medicine

N2CR INVITED SEMINAR

Efficient, High-Resolution Genomic Data Integration Methods for Complex Diseases



Dr Shilpa Garg

Assistant Professor, University of Copenhagen, Denmark

Abstract

Reconstructing the complete phased sequences of every chromosome copy in human and non-human species are important for medical genetics. The unprecedented advancements in sequencing technologies have opened up new avenues to reconstruct these phased sequences that would enable a deeper understanding of molecular, cellular and developmental processes underlying complex diseases including cancer genomics. Despite these interesting sequencing innovations, annotation of novel expression and methylation results are incomplete and inaccurate, and these affect the interpretation of molecular genetics and epigenetics of diseases. There is a pressing need for streamlined, production-level, easy-to-use computational approaches that can reconstruct high-quality chromosome-scale phased sequences which can be applied to hundreds of human genomes.

In this talk, first, I will present an efficient combinatorial phasing model that leverages new long-range strand-specific technology and long reads to generate chromosome-scale phasing. Second, I present an efficient algorithm to perform accurate haplotype-resolved assembly of human individuals. This method takes advantage of new long accurate data type (PacBio HiFi) and long-range Hi-C data. We can, for the first time, generate accurate chromosome-scale phased assemblies with base-level-accuracy of Q50 and continuity of 25Mb within 24 hours per sample, therefore, setting up a milestone in the genomic community. Third, I will present the generalized graph-based method for phased assembly of related individuals. This graph framework provides a compact representation to encode various data types and can be applied to genomes of any complexity having varying heterozygous rates and repeat content specifically for cancer genomics. Finally, I will present the importance of haplotype-resolved assemblies in various medical applications.

These computational tools will be useful to large-scale international initiatives such as the Human Pangenome Reference Consortium and Singapore Diagnostics Development Hub, to yield new insights into unelucidated characteristics of genomes, and enable discoveries in precision medicine of complex diseases such as cancer and Type 2 diabetes.

Biography

Shilpa is an Assistant Professor at University of Copenhagen, Denmark. Prior to this, she was a K99/R00 research fellow at Harvard Medical School and finished her PhD in Computer Science at Max Planck Institute for Informatics. She is passionate to apply the power of computers to understand the biology of complex diseases analysis. She has published in ISMB, ESA, and in prestigious journals like Bioinformatics, Nature Communications and Nature Biotechnology.

July 29 2022, Friday, 4 - 5pm
Online: ZOOM
Host: A/Prof Edward Chow

Scan here to join!