Published on *BSC-CNS* (https://www.bsc.es)

Inicio > Virtual BSC RS/Life Session/BioInfo4Women Seminar: Advanced computational approaches for understanding allele-specific biology of complex diseases

Virtual BSC RS/Life Session/BioInfo4Women Seminar: Advanced computational approaches for understanding allele-specific biology of complex diseases

Objectives

Abstract: Reconstructing the complete phased sequences of every chromosome copy in human and non-human species are important for medical, population and comparative 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. Despite these interesting sequencing innovations, the highly polymorphic and gene-dense regions human leukocyte antigen (HLA) are not yet fully phased in the reference genome. The reference genome still contains gaps in multi-megabase repetitive regions, and thus annotating novel expression and methylation results are incomplete and inaccurate, that affect the interpretation of molecular genetics and epigenetics of diseases. There is a pressing need for a streamlined, production-level, easy-to-use computational approaches that can reconstruct high-quality chromosome-scale phased sequences, and that 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 for the first time can 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 generalised 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. Finally, I will present the importance of haplotype-resolved assemblies to various medical applications including cancer genomics. In summary, my works efficiently and robustly combine data from a variety of sequencing technologies to produce high-quality diploid assemblies. These computational methods will enable high-quality precision medicine and facilitate new and unbiased studies of human (and non-human) haplotype variation in various populations which are currently goals of the Human Genome Reference Project.

K99/R00 Principal Investigator at Harvard Medical School for a year and a half. I am passionate to apply the power of computers to understand the biology of complex diseases. analyses. My work as the first author has been published in ISMB, ESA, and in prestigious journals like Bioinformatics, Nature Communications, Nature Biotechnology, and Genome Biology.

Speakers

Speaker: Shilpa Garg, Tenure-track assistant professor and NNF Data Science Investigator at the University of Copenhagen
Host: Natasa Przulj, BSC Life Sciences - Leading Researcher - Integrative Computational Network Biology Barcelona Supercomputing Center - Centro Nacional de Supercomputación

---