

Keck School of Medicine of USC

Department of Translational Genomics - Virtual Distinguished Lecture Series



Shilpa Garg, PhD
Tenure-track Assistant Professor
University of Copenhagen, Denmark

“Efficient, high-resolution bioinformatic approaches for integrative sequencing analysis of complex diseases”

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. Despite these interesting sequencing innovations, the highly polymorphic and gene-dense 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 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, there-fore, 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. Finally, I will present the importance of haplotype-resolved assemblies to various medical applications.

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 many large-scale human and cancer related projects.

Tuesday, May 25, 2021 11:00am – 12:00pm [Register Here](#)