NanoVar & NanoINSight: Characterization of genomic structural variants and transposable elements using low-depth nanopore sequencing

Asmaa Samy*[1], Cheng Yong Tham[2], and Touati Benoukraf[1,2]

[1]Division of BioMedical Sciences, Faculty of Medicine, Memorial University of Newfoundland – 300 Prince Phillip Dr, St John's NL A1B 3V6, Canada

[2]Cancer Science Institute of Singapore, National University of Singapore – 14 Medical Drive Singapore 11, Singapore

Abstract

Structural variants (SVs) are large genome alterations that are diverse in size and type including unbalanced, balanced, and complex variants. SVs are a major source of genetic diversity and many SVs are associated with Mendelian and complex diseases such as cancer and neurodevelopmental disorder. Our ability to properly characterize SVs was limited till the third-generation sequencing technologies were developed and significantly improved SV discovery despite the high sequencing error rate and low sequencing throughput. Transposable elements (TEs) are a subcategory of SV that play a crucial role in human evolution, genome instability, and disease development. Accurate characterization and mapping of TEs sequences is challenging due to their repetitive nature. Thus, short-read technologies fail to detect TE insertions. In 2020, our research group developed NanoVar, an optimized SV caller utilizing low-depth (8X) long-read sequencing data. NanoVar performs robustly in different genomic studies including mammalian and plant genomes. It outperformed other callers in genotyping accuracy and at low sequencing depths. Here, we present the latest release of NanoVar that includes a new feature called NanoINSight that allows characterization and comprehensive annotation of non-reference repetitive elements. NanoINSight workflow consists of utilizing the novel insertions' sequences extracted by NanoVar to generate multiple sequence alignments, from which consensus sequences are generated and then used as input for RepeatMasker to identify and classify repetitive elements. The updated version of NanoVar demonstrates faster and more precise performance compared with its earlier versions when tested with simulated and real datasets. We anticipate that accurate detection of SVs and novel insertions (TEs) by NanoVar/NanoINSight will contribute to a better understanding of diseases' genesis and development for ultimately improving diagnostic and therapeutic strategies.

Keywords: NanoVar, NanoINSight, structural variants, transposable elements, long, read sequencing, low, depth coverage

*Speaker