Canadian Journal of Biotechnology

Poster Presentation   **OPEN ACCESS**

Category: Bioinformatics

A validated pipeline for detection of SNVs and short InDels from RNA Sequencing

Nitin Mandloi, Rohit Gupta and Ravi Gupta[*]

Research & Development Bioinformatics, MedGenome Labs Pvt. Ltd., Bangalore, INDIA

Presenting author: nitin.m@medgenome.com; [*]Corresponding author: ravig@medgenome.com

Abstract

Effective identification of genomic variations is a crucial step to understand the relationship between genotype and phenotype and it can yield important insights into rare diseases and cancer biology. Exome and whole-genome sequencing are now routinely used in clinics for detection of disease causing variants in Mendelian disorders. Despite great success achieved using these methods, the diagnostic rate for identifying the right variant is ~25-50%. Recent studies have shown that using RNA sequencing (RNA-seq) the diagnostic rate can be improved and can be used as a complementary method for clinical diagnostic.

In this study, we have developed a pipeline to detect germline variants from RNA-seq data. The pipeline steps include: pre-processing, alignment, GATK best practices for RNA-seq and variant filtering. The pre-processing step includes base and adapter trimming and removal of contamination reads from rRNA, tRNA, mitochondrial DNA and repeat regions. The read alignment of the pre-processed reads is performed using STAR/HiSAT. After this we used GATK best practices for the RNA-seq dataset to call germline variants. We benchmarked our pipeline on NA12878 RNA-seq data downloaded from SRA (SRR1258218). After variant calling, the quality passed variants were compared against the gold standard variants provided by GIAB consortium. Of the total ~3.6 million high quality variants reported as gold standard variants for this sample (considering whole genome), our pipeline identified ~58,104 variants to be expressed in RNA-seq. Our pipeline achieved more than 99% of sensitivity in detection of germline variants.