

ABSTRACT

Name : Stefanus Bernard
Study Program : Bioinformatics
Title : Comparison of Bioinformatics Pipeline for Enrichment Illumina Next Generation Sequencing Systems in Detecting SARS-CoV-2 Virus Strains from Yogyakarta and Central Java, Indonesia

Severe Acute Respiratory Syndrome-Coronavirus 2 (SARS-CoV-2) is a newly emerging virus well known as the major cause of worldwide pandemic due to Coronavirus Disease 2019 (COVID-19). Major breakthroughs in the Next Generation Sequencing (NGS) field were elucidated following the first release of a full length SARS-CoV-2 genome in January 10th 2020, with the hope to turn the table against worsening pandemic situation. Previous studies in respiratory viruses characterization require mapping of raw sequences to the human genome in downstream bioinformatics pipeline as part of metagenomic principles. Illumina, as the major player in the NGS arena, took action by releasing guidelines of improved enrichment kits called the Respiratory Virus Oligo Panel (RVOP) based on hybridization capture method capable of capturing targeted respiratory viruses including SARS-CoV-2. Therefore, allowing direct map of raw sequences data to SARS-CoV-2 genome in downstream bioinformatics pipeline. Consequently, two bioinformatics pipelines emerged with no previous studies benchmarked the pipelines. This study focuses on gaining insight and understanding of target enrichment workflow by Illumina through utilization of different bioinformatics pipelines named as 'Fast Pipeline' and 'Normal Pipeline' to SARS-CoV-2 strains isolated from Yogyakarta and Central Java. Overall, both pipelines work well in the characterization of SARS-CoV-2 samples including in the identification of major studied nucleotide substitutions and amino acids mutations. Certain limitations were identified in terms of pipeline algorithm whereas it is highly recommended in the future studies to design a pipeline in an integrated framework, for instance by using NextFlow, a workflow framework to combine all scripts into one fully integrated pipeline.

Keywords: SARS-CoV-2, Next Generation Sequencing, Illumina, Bioinformatics Pipeline