Abstract

Obesity has become one of the biggest public health problems in the world, including in Indonesia (Haththotuwa et al., 2020). One such way to potentially prevent obesity is through a gene prediction method called polygenic risk score (PRS). PRS uses genomic and phenotypic data to determine which gene variants have a higher or lower potential to develop a certain disease (Lewis & Vassos, 2020). Pipeline (also known as a workflow) in bioinformatics refers to the steps of an analysis where the output of a step becomes the input for the next, until it achieves the end result (Leipzig, 2017). As usual with technologies, the tools and methods of a pipeline are ever evolving with the discovery of better algorithms and computational capabilities (Wratten et al., 2019). Combined with the fact that even a simple PRS analysis can contain multiple different softwares, the importance of a standardized and reproducible pipeline becomes an urgent concern. Hence, the aim of this paper is to create a pipeline in Nextflow to calculate obesity PRS for Indonesian patients using the clumping and thresholding method. The pipeline included initial data cleaning, merging, and analysis, PRS calculation using clumping and thresholding method, and result visualizations (which include ROC curve and bell curve for distribution analysis).

Keywords: Obesity; Indonesia; Polygenic Risk Scores; Nextflow; Pipeline Standardization