[Name Placeholder]: A Reproducible Pipeline for Scalable Metabolomics Data Analysis

[Authors Placeholder]

Metabolites are small molecules (molecular weight < 1500Da) acting as middle or end products of cellular metabolism, and metabolomics is a scientific domain of chemical process involving metabolites. The development of modern instruments strengthen the connection between metabolomics research and big data science. A wide variety of approaches and software such as MzMine and XCMS have been developed for computational metabolomics while most of them were designed for application with single desktop and lack of scalability and reproducibility. Nextflow is a pipeline development tool supporting containerization and high performance computing, which improves the scalability and reproducibility of bioinformatics research. The goal of our project is therefore to develop an open-source tool [name placeholder] involving Nextflow to facilitate metabolomics research. [Name placeholder] can be executed on any UNIX-like systems, along with multiple job scheduler. On the other hand, using a single software tool for metabolites detection was reported to have high false positive rate. We thus incorporated multiple metabolites measuring software to [name placeholder] for peak detection in order to enhance its precision. Moreover, [Name placeholder] provides users multiple statistical analysis algorithms including principle component analysis, volcano plots, hierarchical clustering, etc. To facilitate cooperation, MultiQC, which is commonly used to visualize the result of sequence data processing, was adjusted for metabolomics analysis and included in our pipeline to summarize the results.