Research Reproducibility 2020
Educating for Reproducibility: Pathways to Research Integrity
University of Florida, Gainesville, FL, USA

A Highly-Reproducible Workflow for Untargeted Metabolomics Data Processing

Xinsong Du[1], Luran Manfio[1], Alexander Kirpich[2], William R. Hogan[1], Timothy J. Garrett[3], Dominick J. Lemas[1]

[1]*Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida. {xinsongdu, manfiol, hoganwr, djlemas}@ufl.edu*

[2]*Department of Population Health Sciences, School of Public Health, Georgia State University. akirpich@gsu.edu*

[3]*Department of Pathology, Immunology and Laboratory Medicine, College of Medicine, University of Florida. tgarrett@ufl.edu*

ABSTRACT

Background: Reproducibility of untargeted metabolomics data processing remains a challenge.

Objective: The goal of our study is to build a containerized workflow that can improve reproducibility of untargeted metabolomics data processing.

Findings: We used Nextflow, a pipeline development tool supporting containerization, and high performance computing (HiPerGator) to develop a metabolomics data processing workflow. All code and dependencies for the metabolomics pipeline were packaged as Docker containers. Metabolomics data processing was completed using a batch file for MZmine-2.53. Features of the metabolomics pipeline include: 1) analysis ready data sets, 2) estimates on computational resource allocation and 3) formatted reports that provide descriptive statistics and visualizations related to metabolomics data processing. We tested the performance of our metabolic pipeline using a variety of publicly available untargeted metabolomics data sets. Our analysis found that containerized metabolomics workflow provided stable readouts in terms of the number of features identified across metabolomics samples tested.

Conclusion: Our results demonstrate a Nextflow-based framework for untargeted metabolomics data processing has potential to improve reproducibility at the level of data processing.