

Proteomics

Interrogating fractionation and other sources of variability in shotgun proteomes using quality metrics

Marina Kriek,¹²³⁴ Koena Monyai,³ Tandeka U. Magcwebeba,²³⁴ Nelita Du Plessis,²³⁴ Stoyan H. Stoychev,³ David L. Tabb¹²³⁴

¹. SATBBI (South African Tuberculosis Bioinformatics Initiative), Centre for Bioinformatics and Computational Biology, Stellenbosch University

². DST-NRF Centre of Excellence for Biomedical Tuberculosis Research

³. South African Medical Research Council Centre for Tuberculosis Research

⁴. Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa

⁵. Council for Scientific and Industrial Research, Pretoria, South Africa

<https://onlinelibrary.wiley.com/doi/full/10.1002/pmic.201900382>

Abstract

The increasing amount of publicly available proteomics data creates opportunities for data scientists to investigate quality metrics in novel ways. QuaMeter IDFree is used to generate quality metrics from 665 RAW files and 97 WIFF files representing publicly available “shotgun” mass spectrometry datasets. These experiments are selected to represent Mycobacterium tuberculosis lysates, mouse MDSCs, and exosomes derived from human cell lines. Machine learning techniques are demonstrated to detect outliers within experiments and it is shown that quality metrics may be used to distinguish sources of variability among these experiments. In particular, the findings demonstrate that according to nested ANOVA performed on an SDS-PAGE shotgun principal component analysis, runs of fractions from the same gel regions cluster together rather than technical replicates, close temporal proximity, or even biological samples. This indicates that the individual fraction may have had a higher impact on the quality metrics than other factors. In addition, sample type, instrument type, mass analyzer, fragmentation technique, and digestion enzyme are identified as sources of variability. From a quality control perspective, the importance of study design and in particular, the run order, is illustrated in seeking ways to limit the impact of technical variability.