Jul 10, 2024

Quantitative Proteomic Data Analysis

DOI
dx.doi.org/10.17504/protocols.io.kxygxy5pzl8j/v1

Shiyi Wang[1]

[1]Duke University

ASAP Collaborative Rese…

Shiyi Wang
Duke University

OPEN ACCESS

DOI: dx.doi.org/10.17504/protocols.io.kxygxy5pzl8j/v1

Protocol Citation: Shiyi Wang 2024. Quantitative Proteomic Data Analysis. **protocols.io** https://dx.doi.org/10.17504/protocols.io.kxygxy5pzl8j/v1

Protocol status: Working
We use this protocol and it's working

Created: July 10, 2024

Last Modified: July 10, 2024

Protocol Integer ID: 103180

Keywords: ASAPCRN

Funders Acknowledgement:
Aligning Science Across Parkinson's (ASAP) initiative
Grant ID: ASAP-020607

Disclaimer

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to **protocols.io** is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with **protocols.io**, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

Abstract

Quantitative Proteomic Data Analysis

1 **Data Import and Alignment** - Import data from 15 UPLC-MS/MS analyses into Proteome Discoverer 3.0 (Thermo Scientific Inc.). - Exclude conditioning runs but include 3 replicate SPQC samples. - Align individual LCMS data files based on accurate mass and retention time of detected precursor ions using the Minora Feature Detector algorithm.

2 **Relative Peptide Abundance Measurement** - Measure relative peptide abundance based on peak intensities of selected ion chromatograms of aligned features across all runs.

3 **MS/MS Data Search** - Search MS/MS data against the SwissProt M. musculus database, common contaminant/spiked protein database, and reversed-sequence decoys for false discovery rate determination. - Use Sequest with INFERYS to produce fragment ion spectra and perform database searches. - Database search parameters: - Fixed modification: Cys (carbamidomethyl) - Variable modification: Met (oxidation) - Search tolerances: 2ppm precursor and 0.8Da product ion with full trypsin enzyme rules. - Annotate data at a maximum 1% protein false discovery rate using Peptide Validator and Protein FDR Validator nodes in Proteome Discoverer.

4 **Peptide and Protein Homology** - Address peptide homology using razor rules, exclusively assigning a peptide matched to multiple different proteins to the protein with more identified peptides. - Address protein homology by grouping proteins with the same set of peptides and assigning a master protein based on % coverage.

5 **Data Filtering and Normalization** - Apply a filter to remove peptides not measured at least twice across all samples and in at least 50% of the replicates in any single group. - Total intensity normalization: Sum total intensity of all peptides for a sample and normalize across all samples.

6 **Imputation Strategy for Missing Values** - If less than half of the values are missing in a biological group, impute values with an intensity derived from a normal distribution of all values within the same intensity range (20 bins). - If greater than half values are missing for a peptide in a group and peptide intensity is > 5e6, set measured intensity to 0. - Impute all remaining missing values with the lowest 2% of all detected values.

7 **Trimmed-Mean Normalization** - Exclude the top and bottom 10 percent of the signals. - Use the average of the remaining values to normalize across all samples. - Sum peptide intensities belonging to the same protein into a single intensity for analysis.

8 **Technical Reproducibility Assessment** - Calculate the % coefficient of variation (%CV) for each protein across 3 injections of an SPQC pool. - The mean %CV of the SPQC pools should be 11.7%.

9 **Biological + Technical Variability Assessment** - Measure %CVs for each protein across the individual groups, averaging 19.8%.

10 **Initial Statistical Analysis** - Calculate fold-changes between sample groups based on protein expression values. - Perform two-tailed heteroscedastic t-test on log2-transformed data. - Annotate proteins significantly more abundant (Up, fold change >1.5 and p-value <0.05) or less abundant (Down, fold change < -1.5 and p-value <0.05) in a particular genotype or BioID sample group.

11 **Downstream Analysis** - Use Cytoscape (v.3.9.1) for protein interaction networks. - Perform Gene Ontology (GO) enrichment analysis using the ClusterProfiler package for R, with all M. musculus genes as the reference background.