Updated on 09/04/2018

Proteomics data analysis

Mass spectrometer (MS/MS) generated raw files will be used for label-free quantitation (LFQ) of proteins. Base peak chromatograms will be inspected visually using RawMeat, which is a data quality assessment tool designed for Thermo instruments (REF 1). All raw files will be processed together in a single run by MaxQuant (version 1.6.2.3) with default parameters (REF 2). Database searches will be performed using the Andromeda search engine (a peptide search engine based on probabilistic scoring) with the UniProt-SwissProt human canonical database as a reference and a contaminants database of common laboratory contaminants (REF 3).

MaxQuant reports summed intensity for each protein, as well as its iBAQ value. Proteins that share all identified peptides will be combined into a single protein group. Peptides that match multiple protein groups ("razor" peptides) are assigned to the protein group with the most unique peptides. MaxQuant employs the MaxLFQ algorithm for label-free quantitation (LFQ). Quantification will be performed using razor and unique peptides, including those modified by acetylation (protein N-terminal), oxidation (Met) and deamidation (NQ). PTXQC3 and/or Maxreport will be used for general quality control of proteomics data, which takes MaxQuant result files (REF 4, 5).

Data processing will be performed using Perseus (version 1.6.1.3) (REF 6). In brief, protein group LFQ intensities are log2-transformed to reduce the effect of outliers. To overcome the obstacle of missing LFQ values, missing values will be imputed before fit the models. Hierarchical clustering is performed on Z-score normalized, log2-transformed LFQ intensities. Log ratios will be calculated as the difference in average log2 LFQ intensity values between experimental and control groups. Two-tailed, Student's t test calculations will be used in statistical tests. A protein is considered statistically significant if its fold change is ≥ 2 and FDR ≤ 0.1. All the identified differentially expressed/abundant proteins will be used in protein network or pathway analysis.

Required:

1. Raw data files
2. Metadata spreadsheet

Deliverables of proteomics data analysis service:

1. Detailed information about identified protein groups (see additional details at https://www.nature.com/articles/nprot.2016.136/tables/2).
2. Clustering and profile plots of selected clusters.
3. Differential expression analysis with multiple hypothesis testing.
4. Enrichment analysis of protein annotations/pathways.
5. Finally, analysis/methodology description for manuscript

References:

Updated on 09/04/2018

1. http://proteomicsresource.washington.edu/protocols06/RawMeat_1007.exe
2. Cox, J. and Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol, 2008, 26, pp 1367-72.
3. https://www.uniprot.org
4. https://github.com/cbielow/PTXQC
5. http://websdoor.net/bioinfo/maxreport/
6. http://www.perseus-framework.org/

Option 1: Proteomics data analysis by EICC

EICC will analyze your proteomics data using MaxQuant and Persues on Amazon EC2 Instance. Final project cost includes AWS EC2 Computing, Storage (S3/EBS) and EICC analysis fee.

Option 2: Proteomics data analysis by self, with your own AWS account

Get your own AWS account from EICC (with monthly fee $150) and, use EICC's custom AMI designed based on MaxQuant and Persues tools. Final project cost includes AWS EC2 Computing, Storage (S3/EBS) and $150/month.

Note: For proteomics data acquisition please contact Emory Integrated Proteomics Core (EIPC@emory.edu).

Questions? Comments?

Please email us at EICC@emory.edu