Clinical Metaproteomic Data Analysis using Galaxy Training Network

Dechen Bhuming1,2, Katherine Do1,2, Subina Mehta1, Andrew Rajczewski1, Monica Kruk1, James Johnson1, Reid Wagner3, Timothy Griffin1, Pratik Jagtap3

1University of Minnesota-Twin Cities, Department of Biochemistry, Molecular Biology, and Biophysics; 2Life Sciences Summer Undergraduate Research Program

Introduction

Mass spectrometry (MS)-based metaproteomics
- MS-based metaproteomics analyzes proteins expressed by microbial communities (proteomes) in clinical samples, revealing their contributions to disease.
- It characterizes the expressed proteomes of microbiota on a large scale, providing insights into disease response beyond metagenomics which identifies community members. Clinical metaproteomics also uncovers host-microbiome interactions.
- Challenges for microbial peptide detection include interference from abundant host proteins, reduced sensitivity, and potential false positives.

Objective
- We seek to develop training materials to empower others to conduct clinical metaproteomic data analysis, using example data from a metaproteomic study of ovarian cancer (OVCA).
- Our goal was to use the Galaxy Training Network, as part of the Galaxy bioinformatics ecosystem, to develop accessible training materials for clinical metaproteomics analysis.

Methods

Database Generation
- UniProt Knowledgebase (uniprot.org) is a comprehensive database of information about proteins from all living organisms.

Discovery Module
- Microbial peptide identification workflow using three search algorithms -
  - Three search algorithms (SearchGUI/PeptideShaker, MaxQuant and FragPipe) were used to detect microbial peptides from the mass spectrometry data.
  - Complementary search algorithms allow for detection of microbial peptides and a subsequent filtering step to detect microbial peptide sequences with high confidence.

Verification Module
- UniPept (unipept.ugent.be) is a tool that allows for the peptide-centric analysis, particularly in the context of microbial taxonomic communities and functional metaproteomics.

Quantitation Module
- PepQuery provides a peptide-centric approach to proteomic validation of genomic alterations without customized database construction.

Results

- Figure 2. Example Results from the Quantitation Module.
  - Limma-normalization of peptide abundance values allowed for quantitation of differentially abundant peptides in normal, benign, and OVCA samples.
  - Through literature survey, we found 137 taxonomic groups, amounting to ~7.34 million protein sequences.
  - Customized database generated after MetaNovo processing: 4,729 protein sequences.
  - Total microbial peptides detected by program (306 total):
    - SearchGUI/PeptideShaker: 47
    - MaxQuant: 293
  - Total microbial peptides verified by PepQuery: 64
  - Quantified human proteins and microbial proteins after quantitation workflow.

Conclusions and Future Work
- We designed a MS-based Galaxy-driven bioinformatics workflow to process microbial and host proteins. This workflow identifies validated microbial peptide candidates for targeted analysis in individual patient samples.
- Next steps include optimizing the workflow, which includes FragPipe implementation and rerunning analysis to verify the quality of the workflows.
- The workflows optimized here will be published as Clinical Metaproteomics module on the Galaxy Training Network to enable access and use by others.

References


Acknowledgements

Research supported by: NIH HLBI 5R25 HL088728-15, NIH grant 1R01HL133649