

# Clinical Metaproteomic Data Analysis using Galaxy Training Network



Dechen Bhuming<sup>1,2</sup>, Katherine Do<sup>1,2</sup>, Subina Mehta<sup>1</sup>, Andrew Rajczewski<sup>1</sup>, Monica Kruk<sup>1</sup>, James Johnson<sup>1</sup>, Reid Wagner<sup>1</sup>, Timothy Griffin<sup>1</sup>, Pratik Jagtap<sup>1</sup>

<sup>1</sup>University of Minnesota-Twin Cities, Department of Biochemistry, Molecular Biology, and Biophysics; <sup>2</sup>Life 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

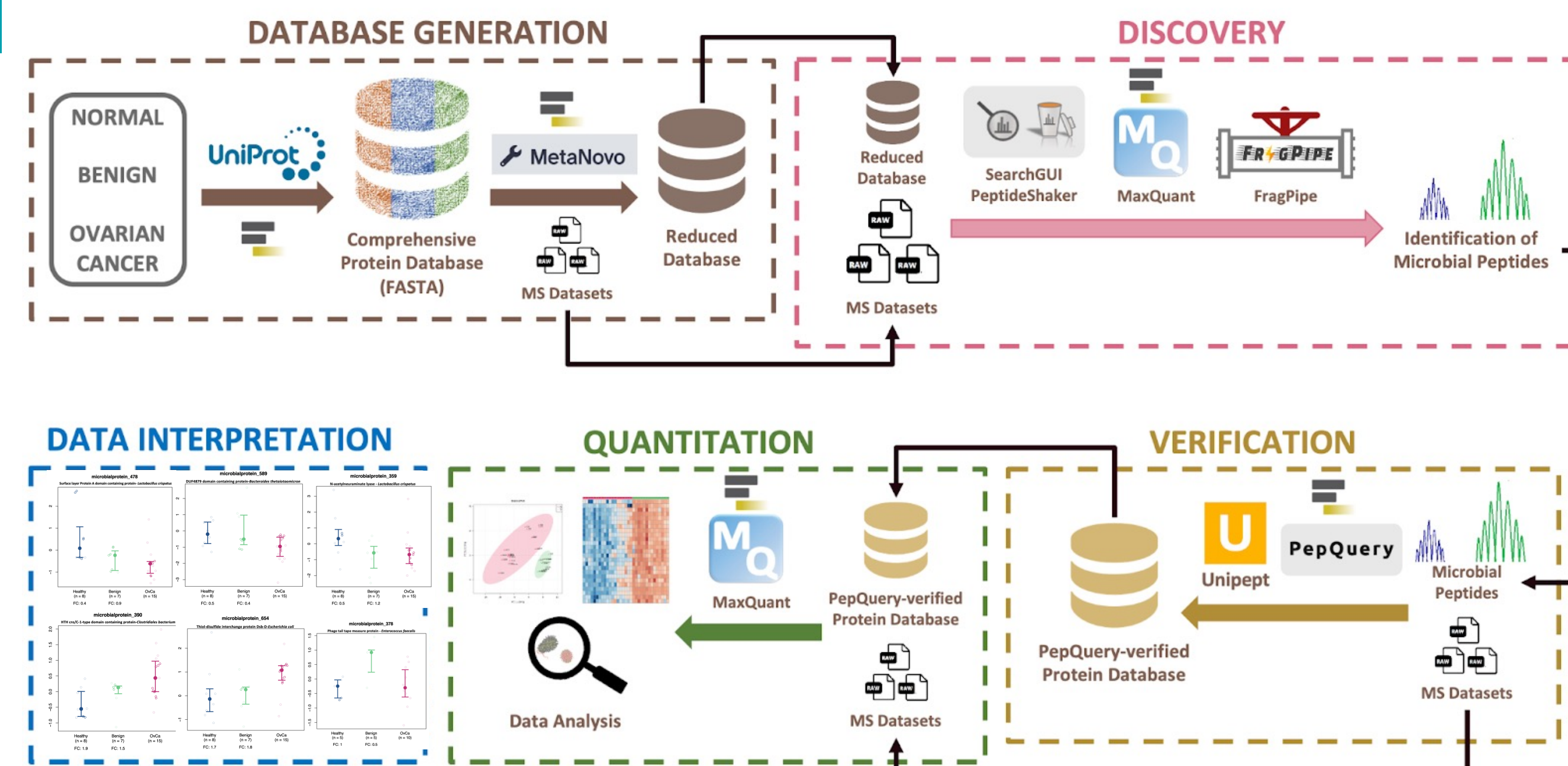
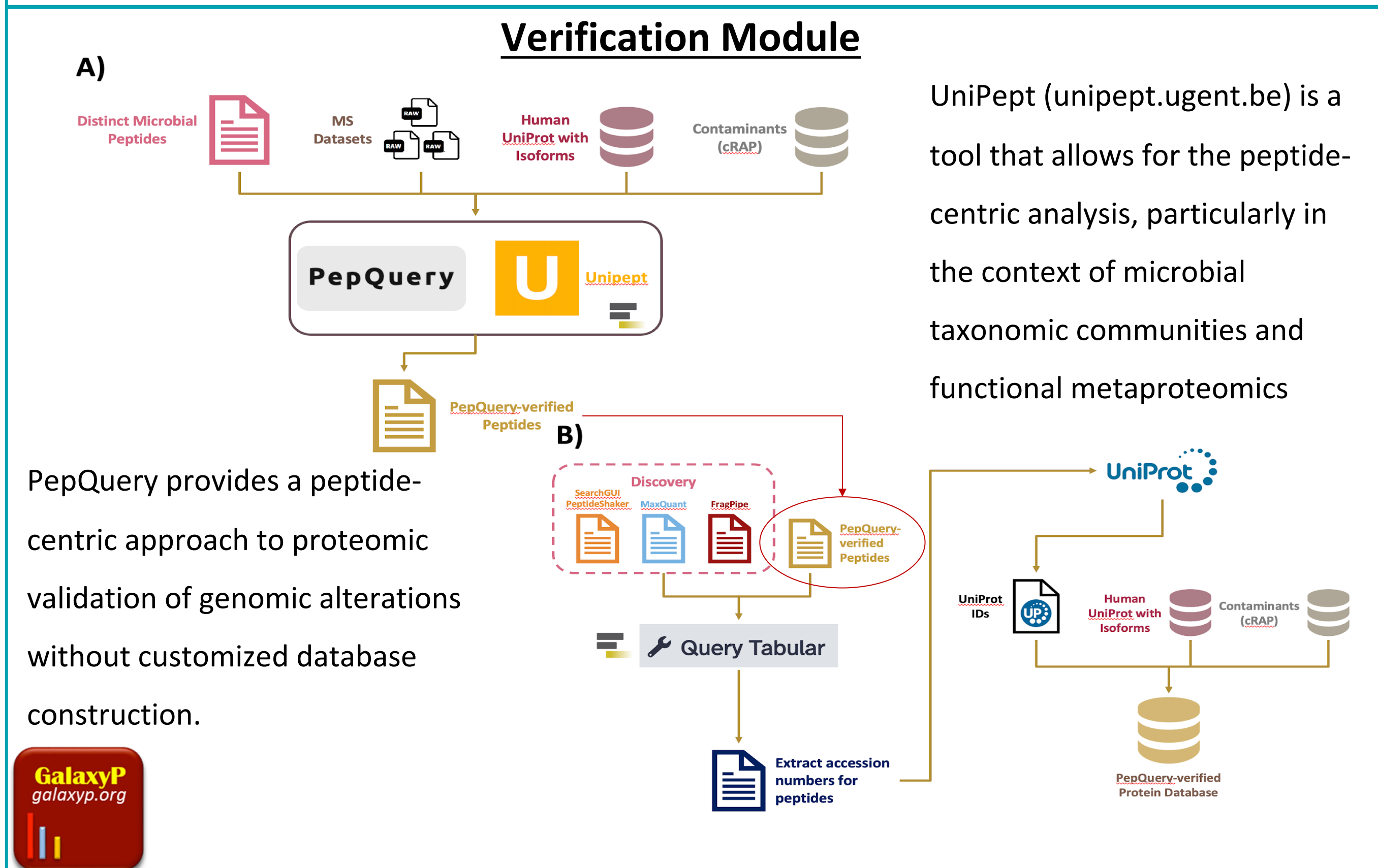
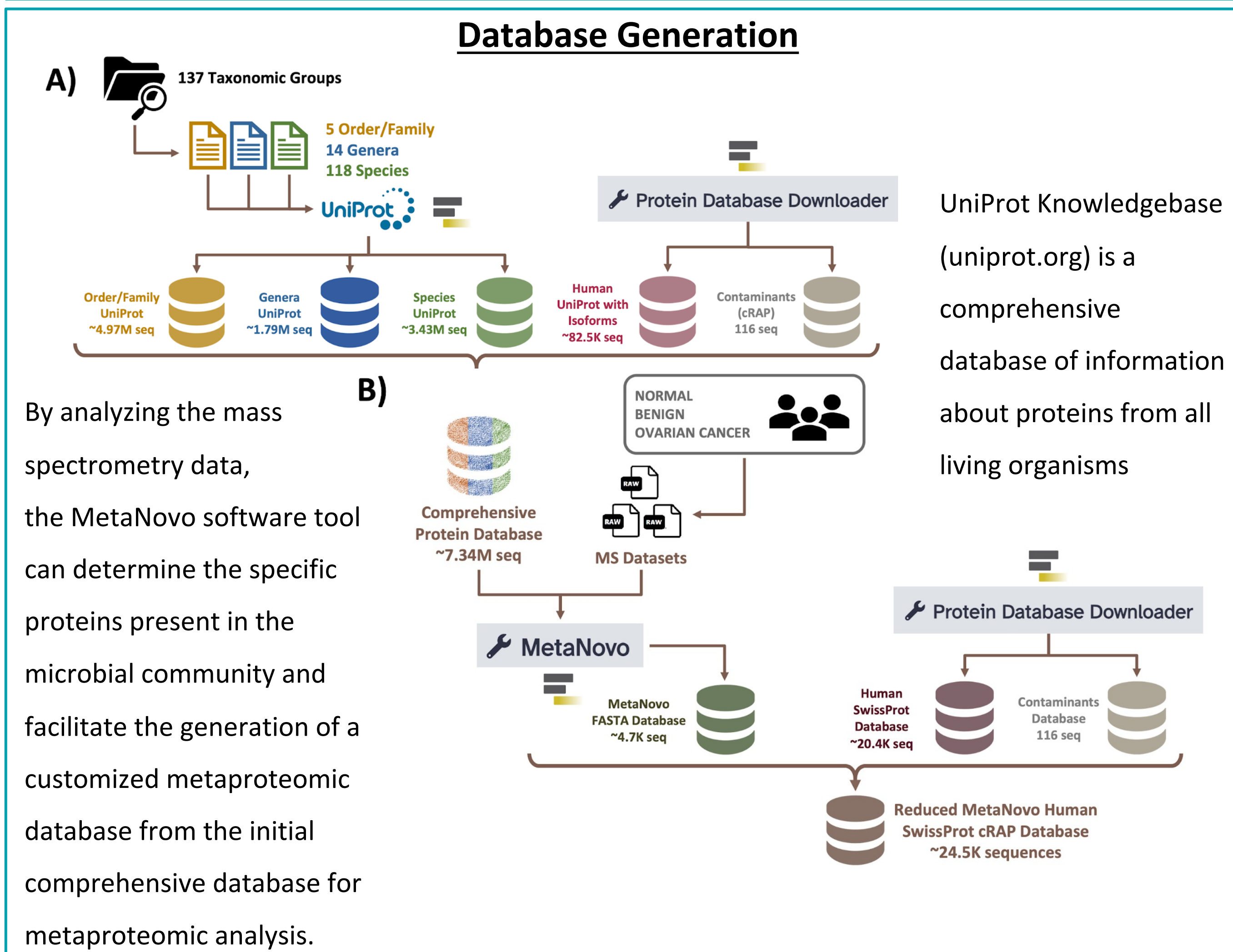
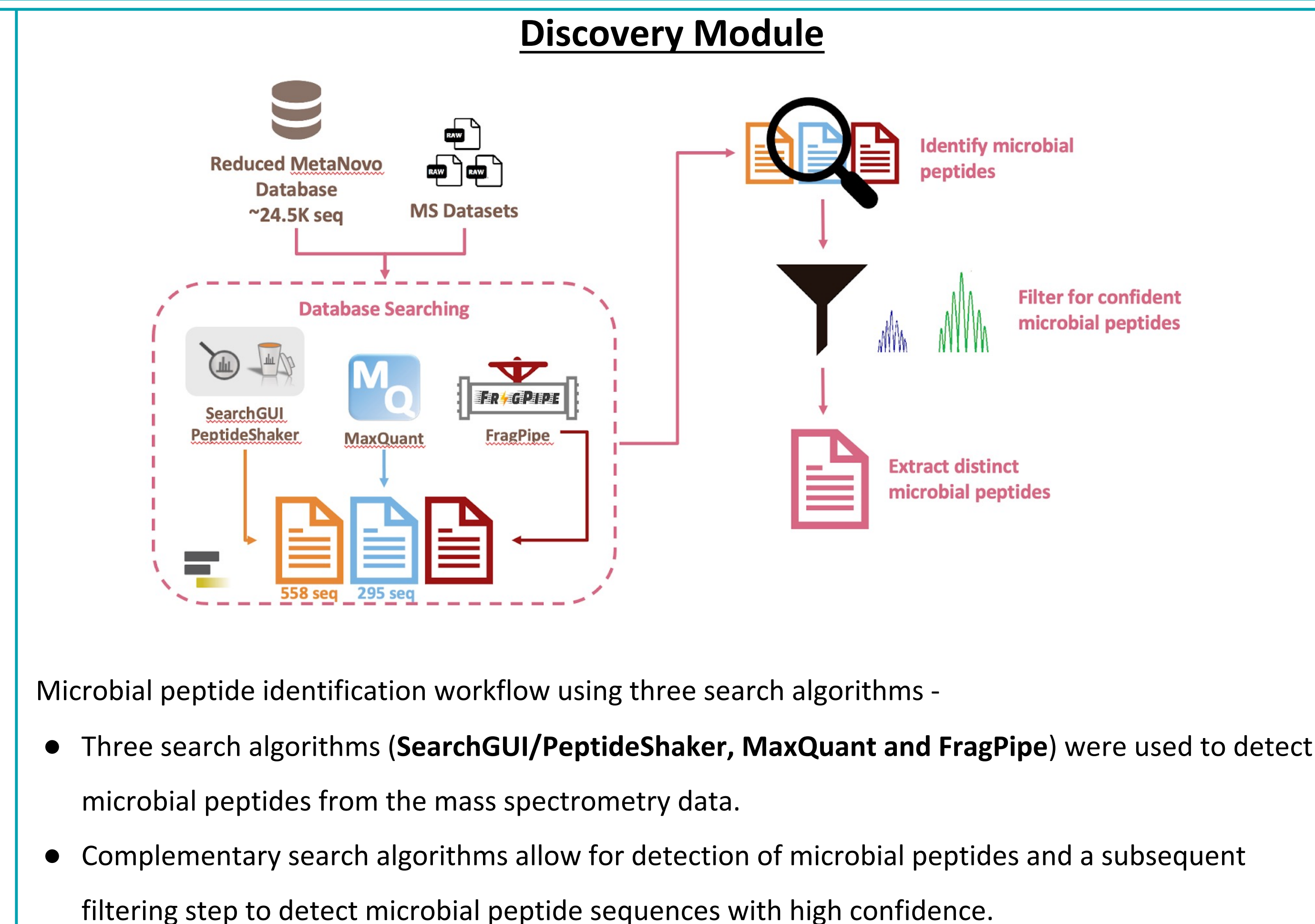


Figure 1. Overview of Galaxy Bioinformatic Workflow for Clinical Metaproteomics.

Figure summarizing the workflow into five modules: a) Database Generation, b) Discovery, c) Verification, d) Quantitation, and e) Data Interpretation.



## Results

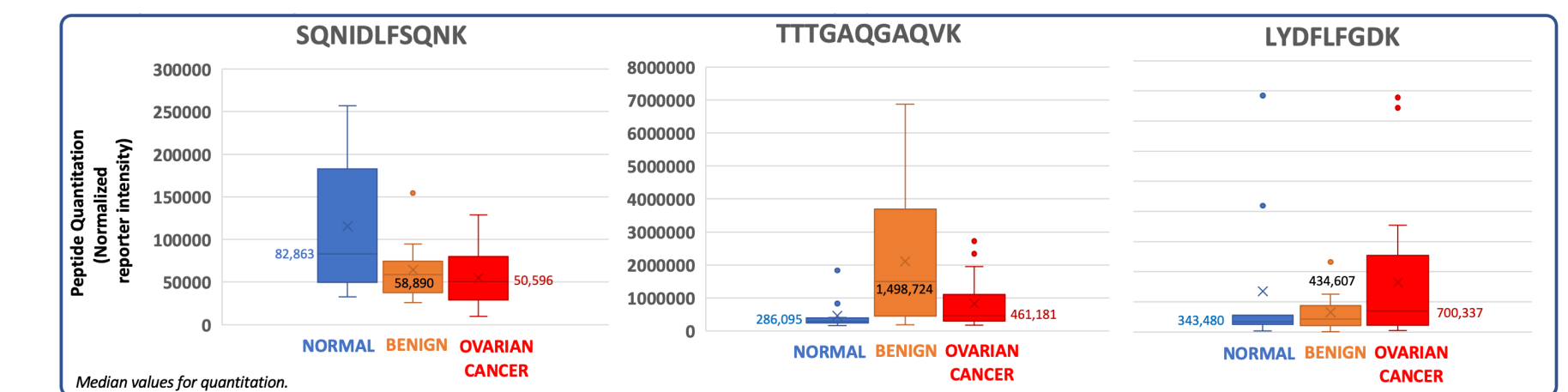


Figure 2. Example Results from the Quantitation Module. Limma-normalization of peptide abundance values allowed for quantitation of differentially abundant peptides in normal, benign, and OVCA samples.

- 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

- Barnes, H., & Vaudel, M. (2018). SearchGUI: A Highly Adaptable Common Interface for Proteomics Search and de Novo Engines. *Journal of proteome research*, 17(7), 2552–2555. <https://doi.org/10.1021/acs.jproteome.8b00175>
- Batut B, Hiltmann S, Bagnacani A, et al. Community-Driven Data Analysis Training for Biology. *Cell Syst*. 2018;6(6):752-758.e1. doi:10.1016/j.cels.2018.05.012
- Blank C, Easterly C, Gruening B, et al. Disseminating Metaproteomic Informatics Capabilities and Knowledge Using the Galaxy-P Framework. *Proteomes*. 2018;6(1):E7. doi:10.3390/proteomes6010007
- Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res*. Published online April 21, 2022:gkac247. doi:10.1093/nar/gkac247
- Timothy J. Griffin, Pratik Jagtap, James Johnson, Clemens Blank, Subina Mehta, 2021 Metaproteomics tutorial (Galaxy Training Materials). <https://training.galaxyproject.org/archive/2022-01-01/topics/proteomics/tutorials/metaproteomics/tutorial.html>
- Tyanova, S., Temu, T., and Cox, J., The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, *Nat Protocols*, 2016, 11, pp 2301–2319
- Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., Martens, L., & Barsnes, H. (2015). PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nature biotechnology*, 33(1), 22–24. <https://doi.org/10.1038/nbt.3109>
- Wen, Bo, and Bing Zhang. "PepQuery2 democratizes public MS proteomics data for rapid peptide searching." *Nature communications* 14.1 (2023): 2213.

## Acknowledgements

Research supported by: NIH HLBI 5R25 HL088728-15, NIH grant 1R01HL136499

