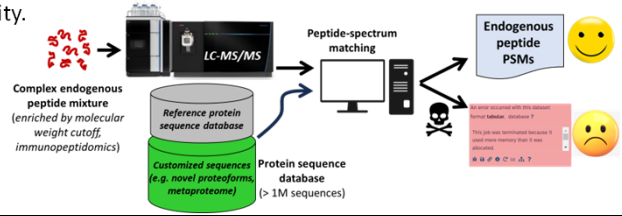
No enzyme? No problem! An accessible, scalable cloud-based solution for efficient MS-based endogenous peptidomics and immunopeptidomics

Reid Wagner[1], James Johnson[1], Subina Mehta[2], Kyle Richards[3], Suzanne Coleman[3], David Largaespada[3], Monica Kruk[2], Chris Wendt[4], Bing Zhang[5], Fengchao Yu[6], Alexey I. Nesvizhskii[6], Pratik Jagtap[2], and Timothy J. Griffin[2]

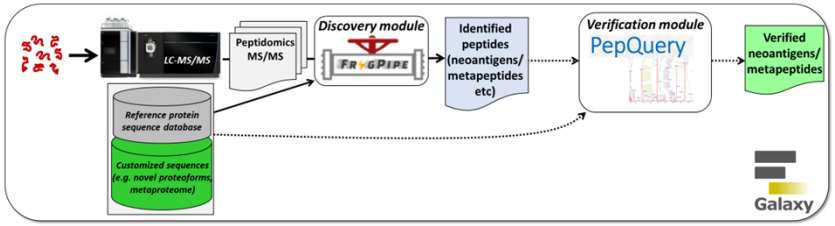
[1]Minnesota Supercomputing Institute, U. of Minnesota; [2]Department of Biochemistry, Molecular Biology and Biophysics, U. of Minnesota; [3]Department of Pediatrics, U. of Minnesota; [4]Minneapolis VA Health Care System and U. of Minnesota Medical School; [5]Department of Molecular and Human Genetics, Baylor College of Medicine; [6]Department of Pathology and Department of Computational Medicine and Bioinformatics, U. of Michigan

BACKGROUND AND RATIONALE

MS-based endogenous peptidomics identifies protease-derived peptides, as well as disease-specific peptide antigens isolated by immunopeptidomics. These "no enzyme" analyses for peptide spectrum match (PSM) generation use large protein sequence database search space requiring memory and processor power exceeding the capabilities of most software platforms. PSMs from non-tryptic peptides require stringent verification, due to unique fragmentation characteristics and, for immunopeptidomic studies, potential matches to putative non-normal neoantigen sequences. We have developed an informatic solution for diverse applications in peptidomics, integrating the FragPipe and PepQuery platforms within the Galaxy bioinformatics ecosystem. We demonstrate rapid and confident peptide identification and verification, including neoantigen sequences and non-host microbial peptides, from clinical samples using large databases (>1M sequences). This unique computational resource offers a critically-needed peptidomics/immunopeptidomics informatics solution available to the research community.



PEPTIDOMICS WORKFLOW



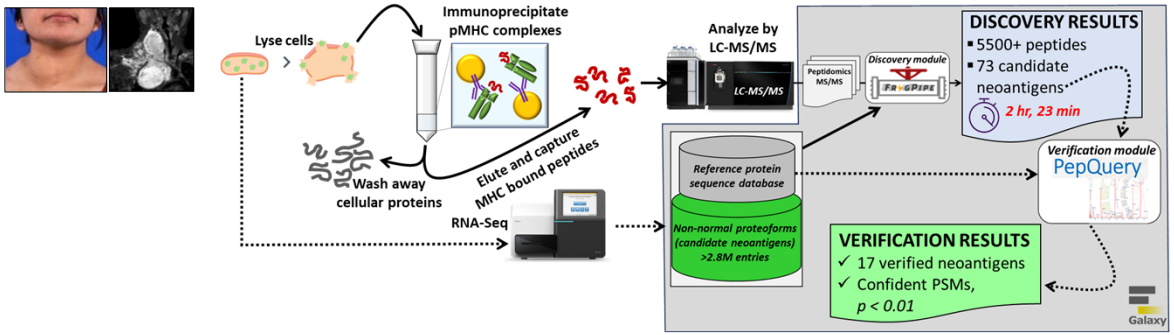
The analysis workflow is comprised of these modules and functionalities:

- ***Customized protein sequence database generation.*** Using specialized Galaxy-based tools, customized protein sequence databases can be created, including those with non-normal proteoforms predicted from transcriptomics data (1) or metaproteome sequences based on taxa present in the sample (2)
- ***Discovery Module.*** This module makes use of the FragPipe suite of tools (3), notably MSFragger for efficiently generating peptide spectral matches (PSMs) using no enzyme settings
- ***Verification Module.*** If desired, PSMs from the Discovery Module can be verified using the PepQuery2 software (4), which rigorously re-evaluates candidate PSMs considering other sequences, including PTMs, as alternative matches; this is particularly important for peptides from novel proteoforms (such as neoantigens) or matches to microbial peptides found within host organisms
- ***Deployment in the Galaxy Ecosystem.*** Tools are available through the Galaxy Tool Shed, accessible on public gateways implemented on high performance computing resources, with options for developing training material via the Galaxy Training Network

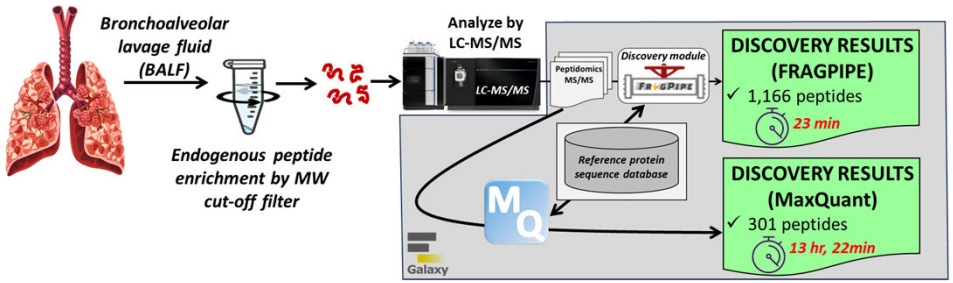
Galaxy Tool Shed · useGalaxy.* · Galaxy Training Network

RESULTS: APPLICATION TO DIVERSE STUDIES UTILIZING PEPTIDOMICS

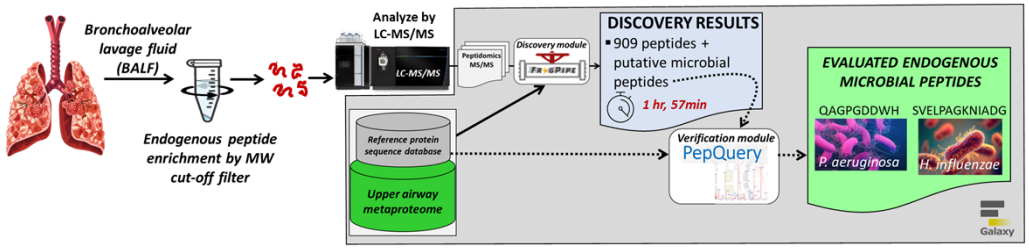
Application 1: Immunopeptidomics in malignant peripheral nerve sheath tumor (MPNST) cells



Application 2: Peptidomics of bronchoalveolar lavage fluid in chronic obstructive pulmonary disease



Application 3: Metapeptidomics of bronchoalveolar lavage fluid in chronic obstructive pulmonary disease



CONCLUSIONS AND FUTURE DIRECTIONS

- FragPipe in Galaxy offers efficient and sensitive peptidomic Discovery analysis
- PepQuery can verify no enzyme PSMS; using narrow peptide length saves on memory requirements while still providing statistics for assessing confidence
- Tools and workflows will be made available along with training material within the Galaxy ecosystem

References

1. An Accessible Proteogenomics Informatics Resource for Cancer Researchers. Chambers MC, et al *Cancer Res.* 2017 77:e43-e46.
2. Metaproteomic analysis using the Galaxy framework. Jagtap PD, et al *Proteomics*. 2015, 15:3553.
3. IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs. Yu F, Haynes SE, Nesvizhskii AI. *Mol Cell Proteomics*. 2021, 20:100077.
4. PepQuery2 democratizes public MS proteomics data for rapid peptide searching. Wen B, Zhang B. *Nat Commun.* 2023, 14:2213.

Acknowledgements

We thank the Minnesota Supercomputing Institute for maintenance of computational infrastructure used for development and testing, and the Center for Metabolomics and Proteomics at the University of Minnesota for support generating peptidomic MS/MS data. This work is supported by NIH grants UG3CA244687, R01HL140971 and 1U01CA288888, as well as grant P30CA077598 to the Masonic Cancer Center at the University of Minnesota

The authors declare no competing financial interest.