# Mass Spectrometry (MS)-based multi-omics analysis using the Galaxy-P bioinformatics platform: A case study in COVID-19 data analysis

*Great Lakes Bioinformatics Conference*
*May 10, 2021*

**Galaxy for Proteomics (Galaxy-P) Team**
**University of Minnesota**

*Learn more at galaxyp.org*
z.umn.edu/itcrgalaxyvideo

# Overview and objectives

- Introduce the Galaxy framework as a solution for data analysis across ‘omics’ domains, with a focus on MS-based proteomics
- Provide **hands-on** experience to attendees in using Galaxy
- Demonstrate use of Galaxy for MS-based proteomics and multi-omic analysis using COVID-19 studies as representative data
- Lay the foundation for attendees to implement Galaxy in their work, and/or contribute tools to the Galaxy community

GalaxyP
galaxyp.org

# Workshop instructors and acknowledgements

- **Instructors**
  - Pratik Jagtap
  - Subina Mehta
  - Andrew Rajczewski
  - An Nguyen

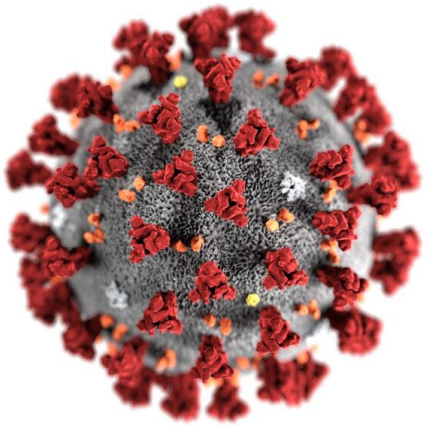
- **Other contributors**
  - James Johnson
  - Reid Wagner
  - usegalaxy.eu team
  - Galaxy community

- **GLBIO Conference organizers**

NATIONAL CANCER INSTITUTE
Informatics Technology for Cancer Research

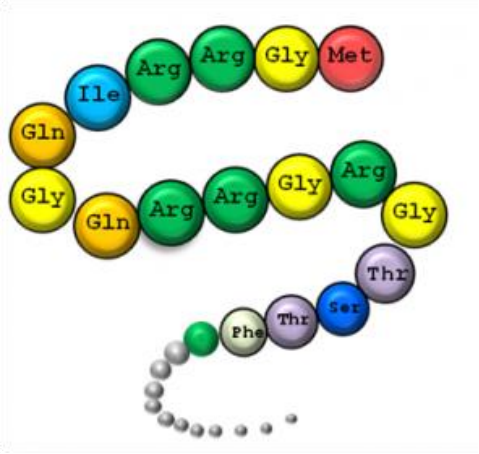
# Topic(s) of workshop

*Mass spectrometry-based proteomic informatics*



*COVID-19 research*

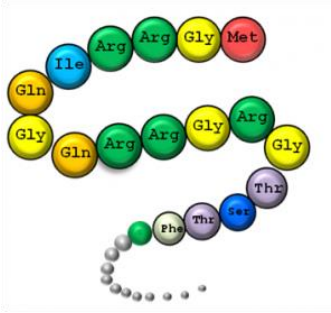
*Metaproteomics (multi-omics)*




*Protein sequence analysis*

GalaxyP
galaxyp.org

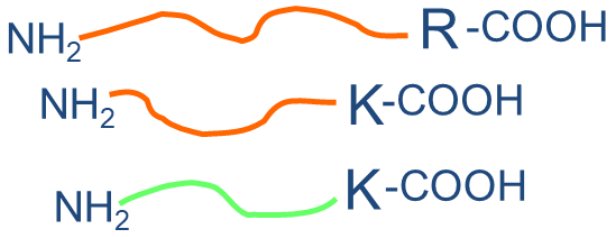
# Shotgun proteomics: breaking proteins into peptides




1. Protein, MW = 10,000 +

digest into peptides (enzymatic, chemical)

2. Peptides, MW < 4,000




## Trypsin

N-TH**K**.NCPHIVVGTPG**R**.IPD-C

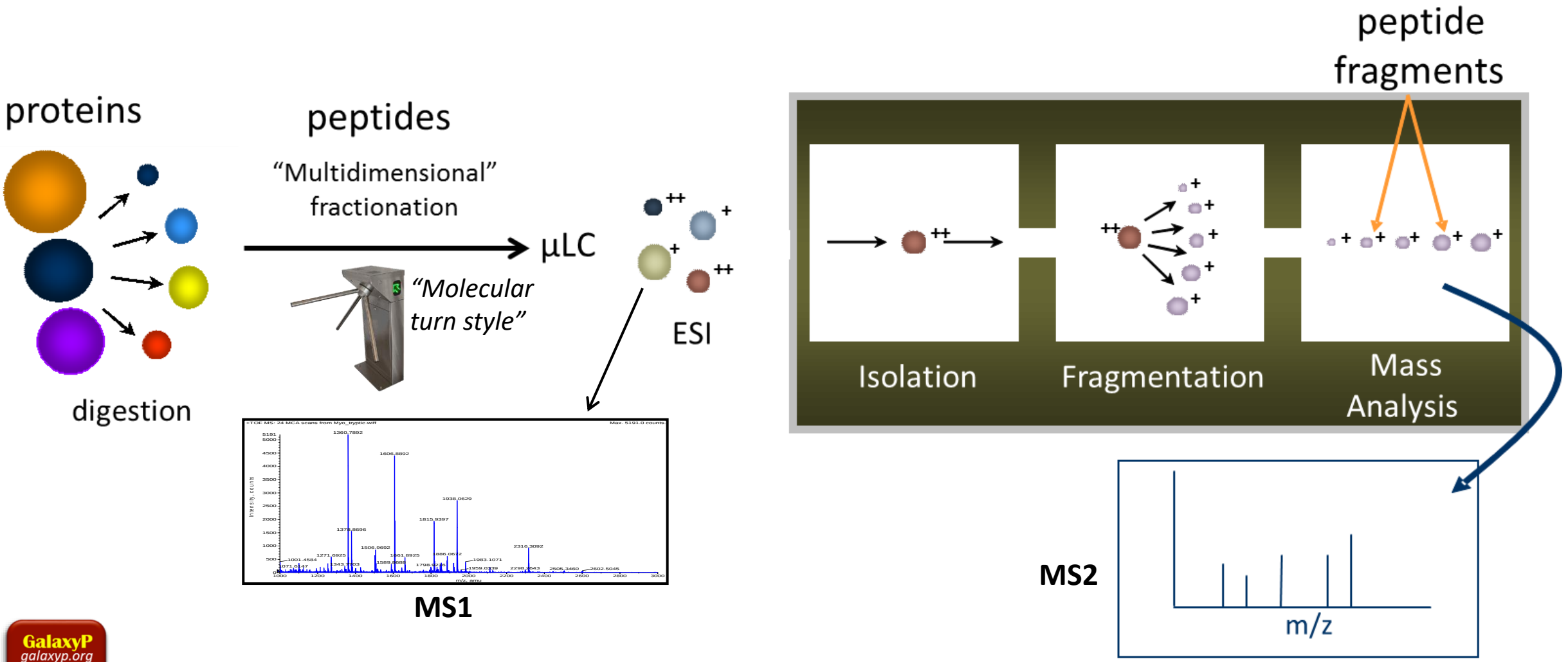
cleaves c-teminal side of arginine (R) and lysine (K)

3. Peptide fragments

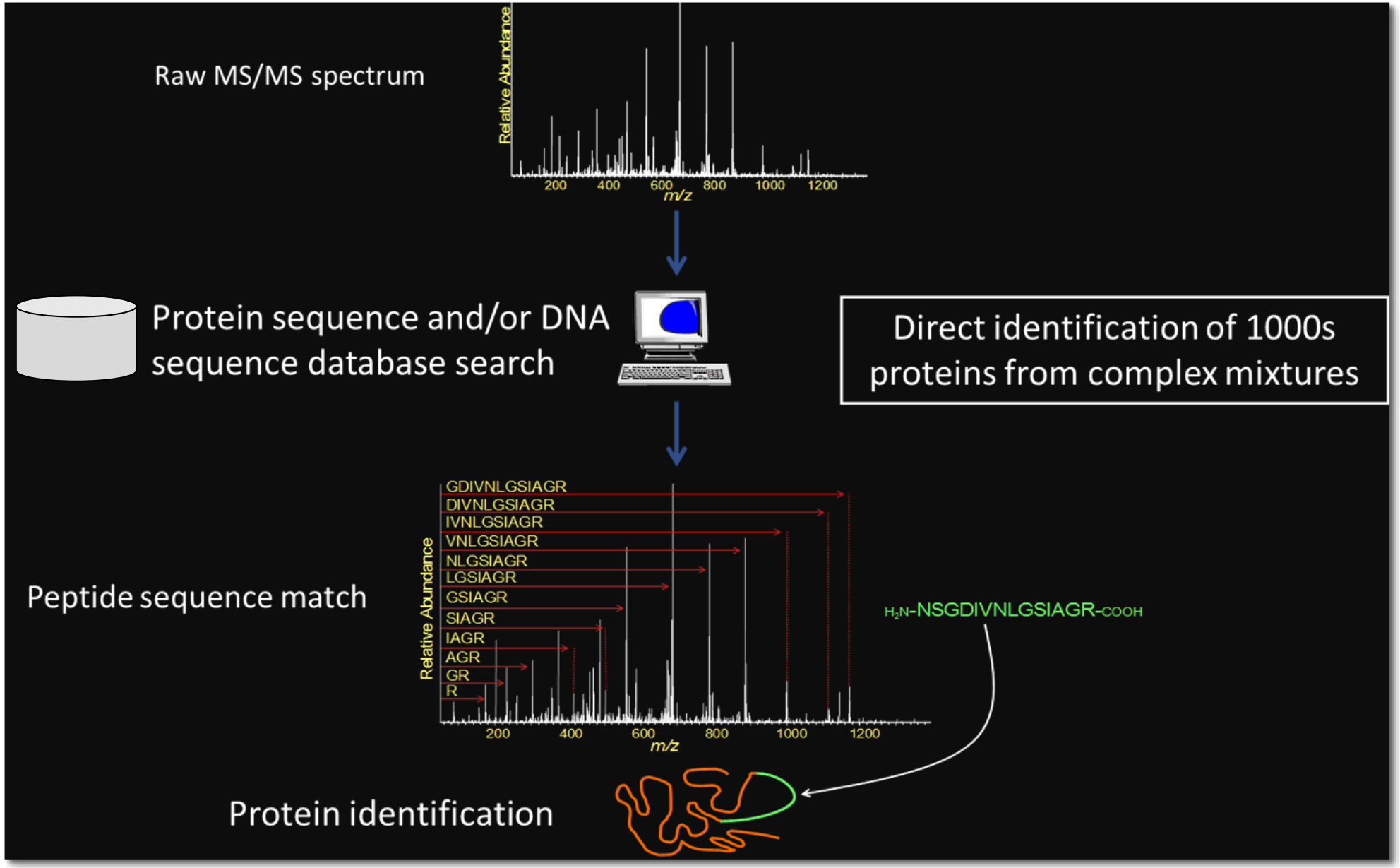
GalaxyP
galaxyp.org

# Mass Spectrometry-based proteomics: A primer

## Peptide fractionation coupled to tandem mass spectrometry (MS/MS)

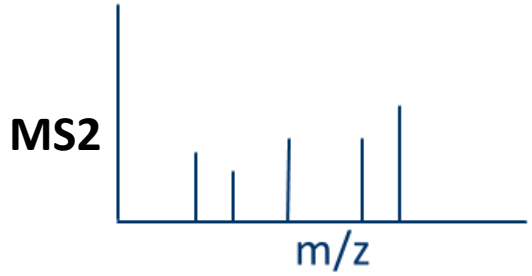


# Matching amino acid sequences to MS/MS data




GalaxyP
galaxyp.org

# Inferring protein identity for peptide sequences




CAQCHTVEK

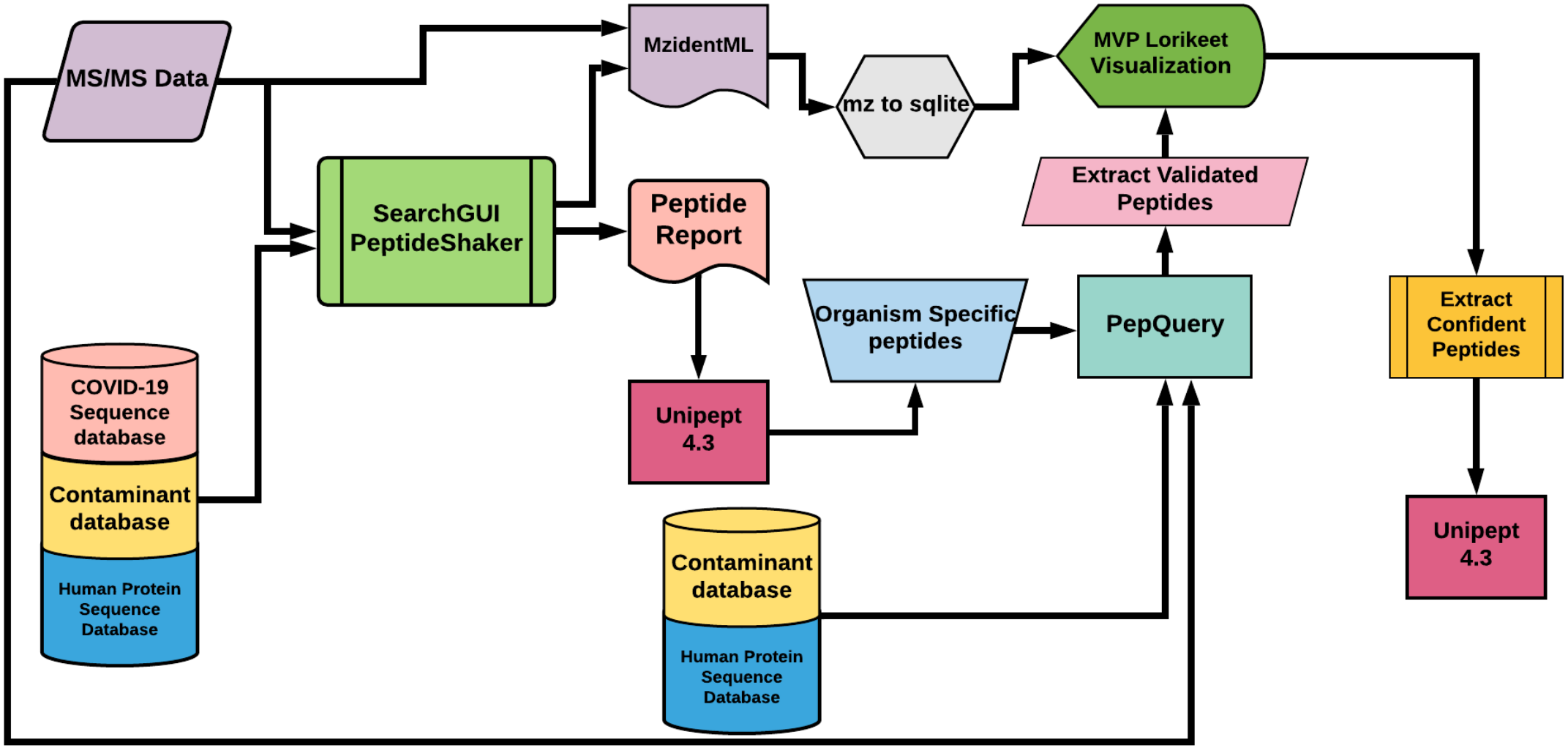
**Cytochrome C**

$NH_2$GDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGFTYTDANKNKGITW
KEETLMEYLENPKKYIPGTKMIFAGIKKKTEREDLIAYLKKATNE$COOH$

GalaxyP
galaxyp.org

# Driven by bioinformatics

- Many step workflow for analyzing large and diverse datasets




GalaxyP
galaxyp.org

# One solution: Galaxy

Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team.. ***Genome Biol.*** 2010, **11**: R86.

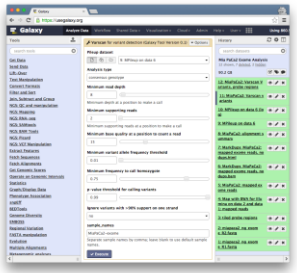
- ✓ A web-based, community developed bioinformatics workbench for integrating disparate software -- flexible
- ✓ Geared towards use by bench scientists; many training resources available
- ✓ Already home to genomic/transcriptomic tools
- ✓ Provenance tracking, sharing and reproducibility
- ✓ Amenable to other 'omic tools (e.g. Galaxy for proteomics project, Galaxy-P)

***Working philosophy:***

# Galaxy: an integrative workbench well-suited for multi-omics

## Interfaces

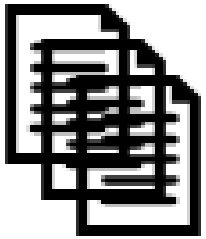
Web UI



Programmatic API

***Integrate datasets, analysis tools, visualizations, and computing resources for large-scale biomedical data science***

## Datasets



## Computing Resources

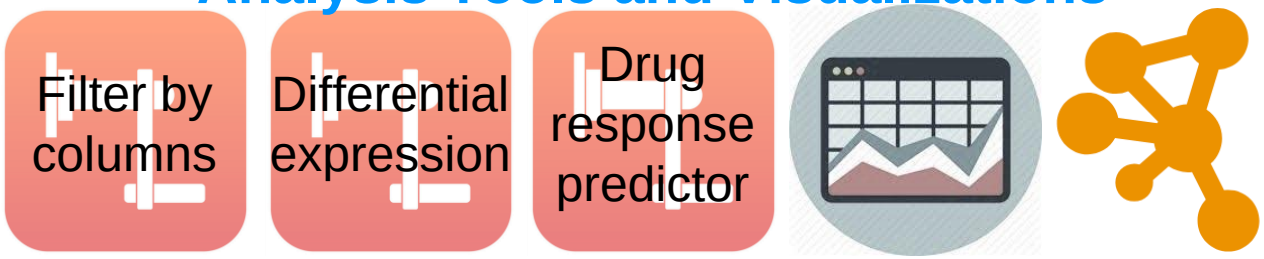
slurm workload manager

Condor High Throughput Computing

Google Cloud Platform

amazon web services

## Analysis Tools and Visualizations




GalaxyP
galaxyp.org

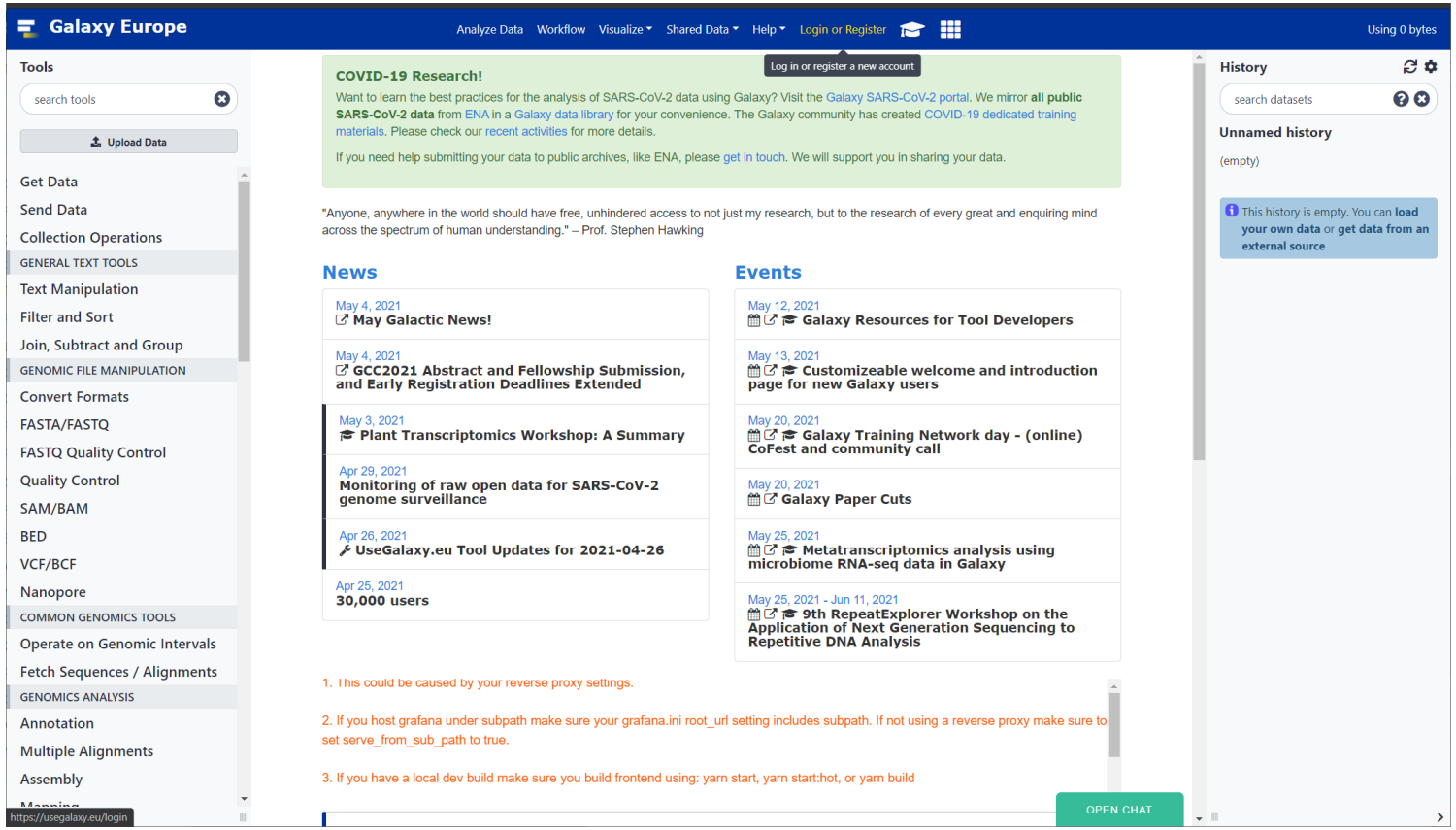
*Courtesy Jeremy Goecks, OHSU*

# Accessing Galaxy via a public gateway

1. Login/Register: [usegalaxy.eu](usegalaxy.eu)

2. Go to TIaaS link: [https://usegalaxy.eu/join-training/glbiotraining](https://usegalaxy.eu/join-training/glbiotraining)

3. Return to [usegalaxy.eu](usegalaxy.eu) site



# The Galaxy Interface

Main viewing window
(workflow development, results visualization etc)

Tools

History

GalaxyP
galaxyp.org

# Built for the community: Tool sharing and access

Galaxy Tool Shed

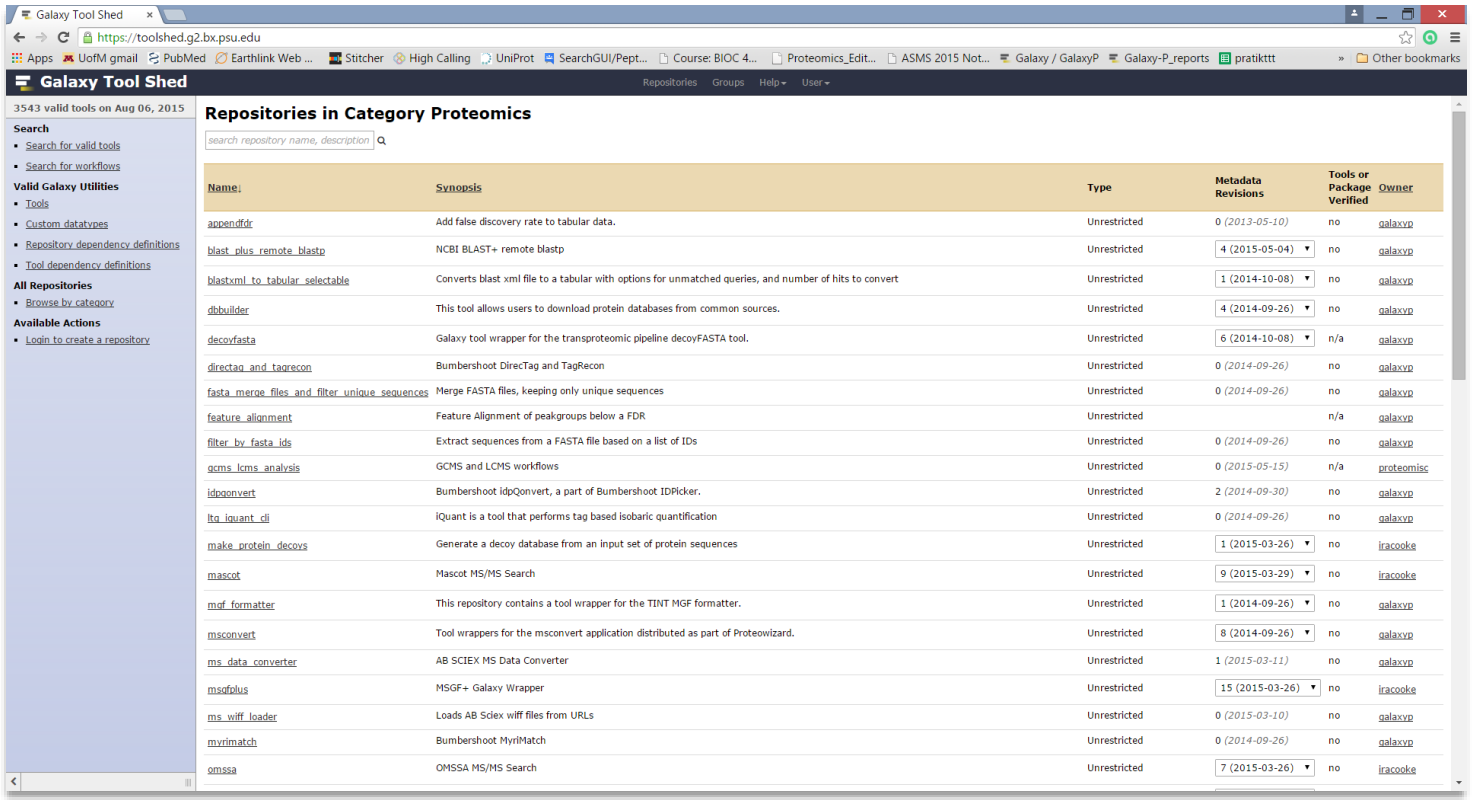
Blankenberg *et al. Genome Biology* 2014, **15**:403
http://genomebiology.com/2014/15/2/403

Genome Biology

OPEN LETTER

## Dissemination of scientific software with Galaxy ToolShed

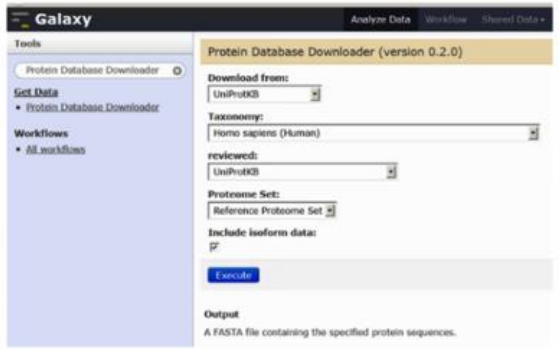
Daniel Blankenberg[1,4], Gregory Von Kuster[1,4], Emil Bouvier[1,4], Dannon Baker[2,4], Enis Afgan[4,5], Nicholas Stoler[3], the Galaxy Team[4], James Taylor[2,4*] and Anton Nekrutenko[1,4*]



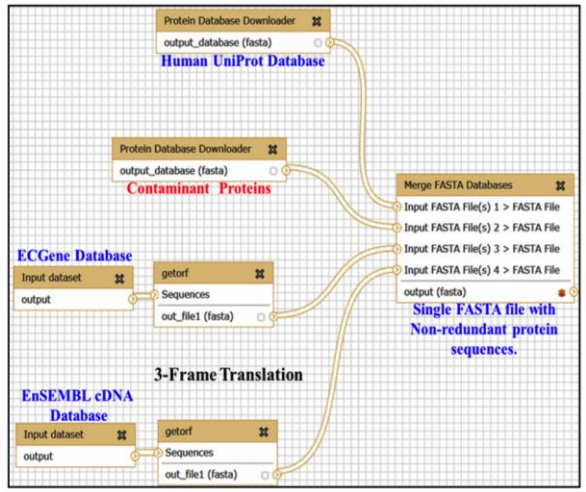

GalaxyP
galaxyp.org

# Defining Workflows and Histories

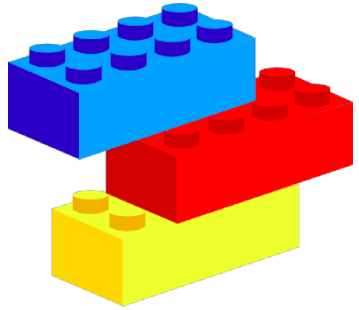
**Workflows:** Multiple software tools used in a sequential manner for an analysis



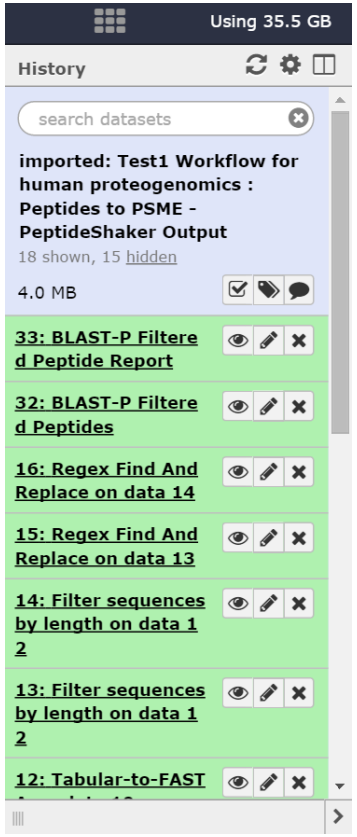

Single software tool in Galaxy




Integrated workflow using multiple, connected tools

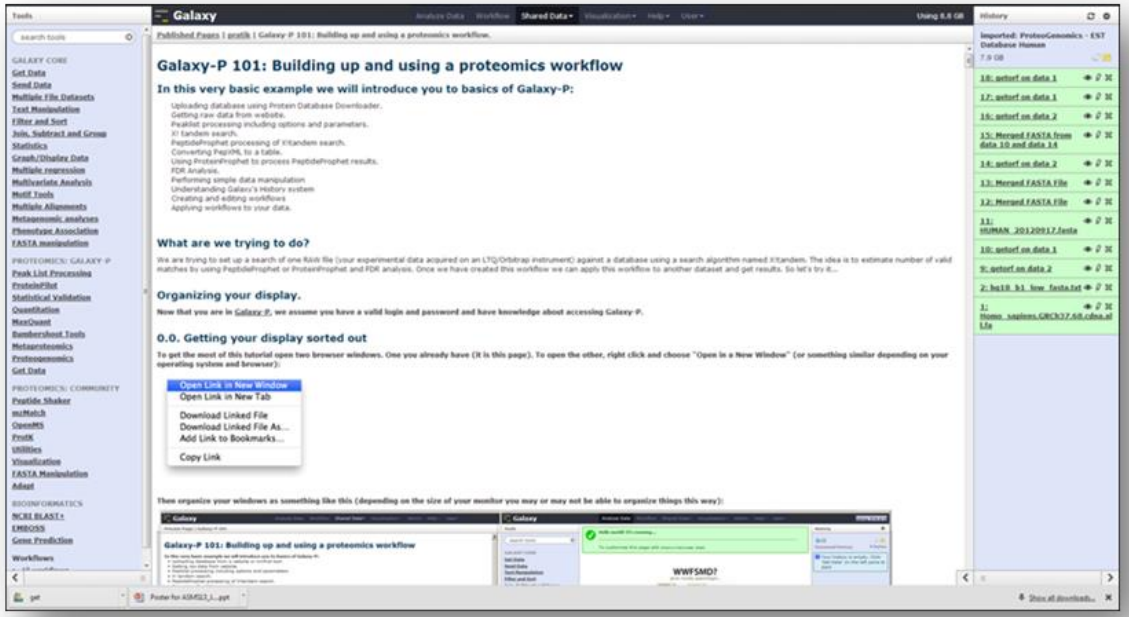


**History:** recorded, complete analysis (workflow + all data and results)




GalaxyP
galaxyp.org

# Built for the community: Sharing workflows and histories




URL export

File export (.ga)

**HISTORY:** https://galaxyp.msi.umn.edu/u/pjagtap/h/itraq-search-yang-2-xtandem-scaffold

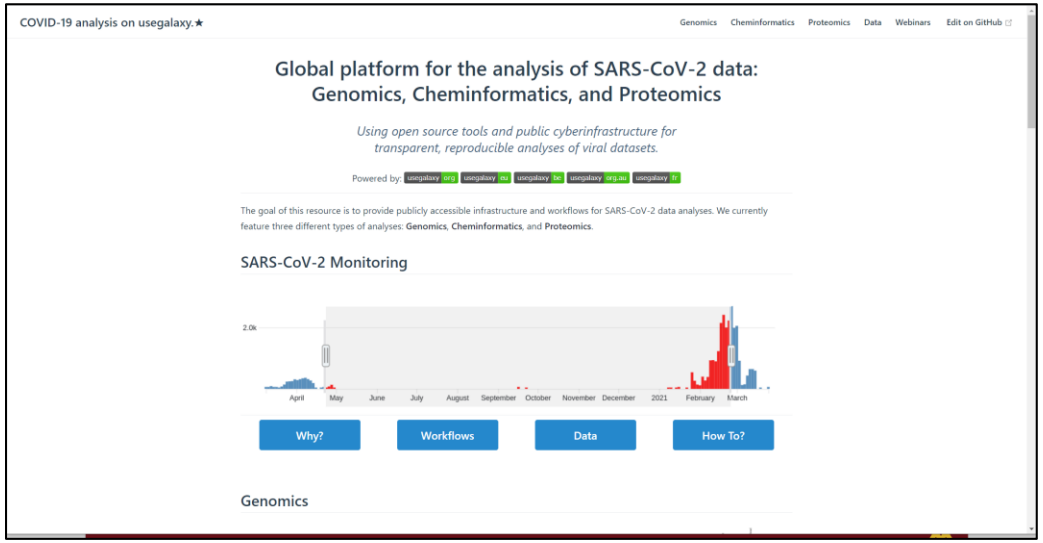
**WORKFLOW:** https://galaxyp.msi.umn.edu/u/pjagtap/w/workflow-for-4-plex-itraq-xtandem-search-scaffold-processing

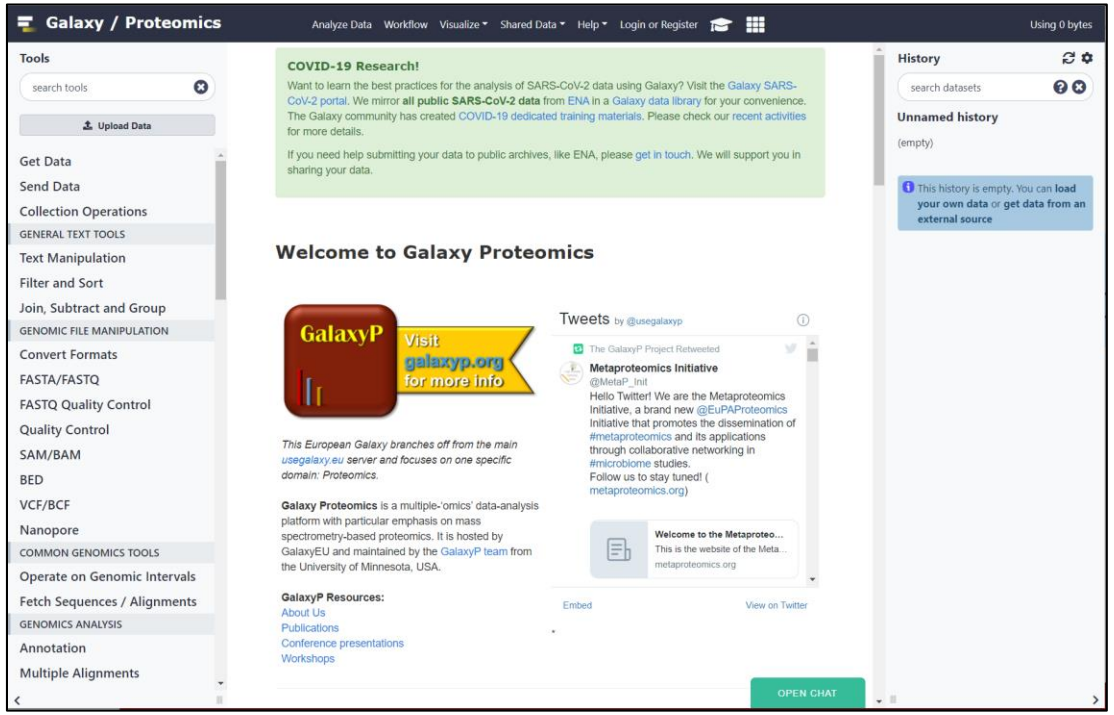
Galaxy-Workflow-Workflow_for_4-plex_iTRAQ_X_tandem_Search_Scaffold_Processing.**ga**

GalaxyP
galaxyp.org

# Accessing and using Galaxy

useGalaxy.*




Galaxy Tool Shed

# Agenda and schedule

**1 PM - 1:30 PM: Introduction to Galaxy and Galaxy-P**

**1:30 PM - 2:00 PM: Mass Spectrometry, Multi-omics and Galaxy Workflows for COVID-19 Research**

**2:00 PM - 3:30 PM: Detecting SARS-CoV-2 and microbial peptides from patient samples**

**3:30 PM - 4:00 PM: Summary and Questions and Answers**

***Please ask questions as they come up!***

GalaxyP
galaxyp.org

# Galaxy Workflows for analysis of COVID-19 Mass Spectrometry datasets

Pratik Jagtap

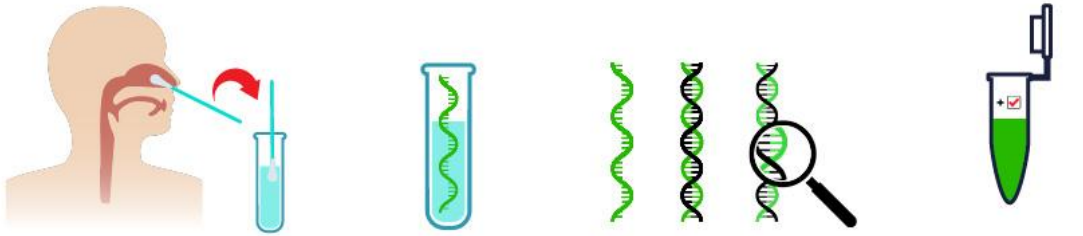
*University of Minnesota*

*Galaxy-P Team*

# COVID-19 DETECTION METHODS

## Molecular Tests (Nucleic Acid Detection)

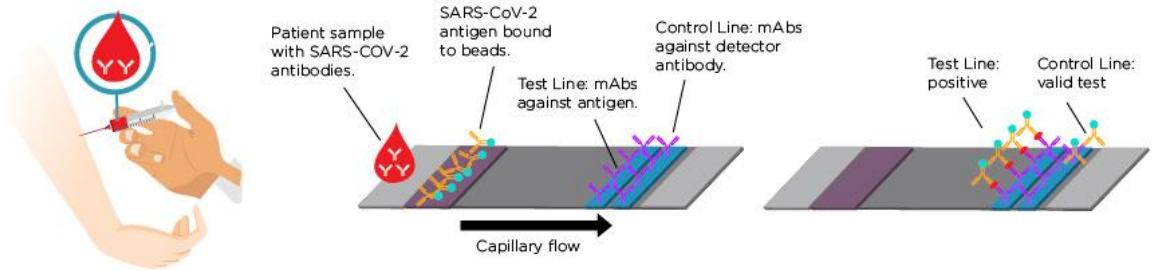
*Diagnose active SARS-CoV-2 infections*



1. Obtain Specimen: Swab.
2. Extract RNA from specimen and convert to DNA.
3. Amplify by PCR with SARS-CoV-2 specific primers.
4. Interpret results: presence of viral RNA indicates active SARS-CoV-2 infection.

## Antibody Tests (Serology)

*Detect immune response to SARS-CoV-2 exposure*




1. Obtain Specimen: Blood Sample.
2. Expose specimen to SARS-CoV-2 specific antigens.
3. Interpret results: color change indicates previous exposure to SARS-CoV-2.

AMERICAN SOCIETY FOR MICROBIOLOGY

asms.org

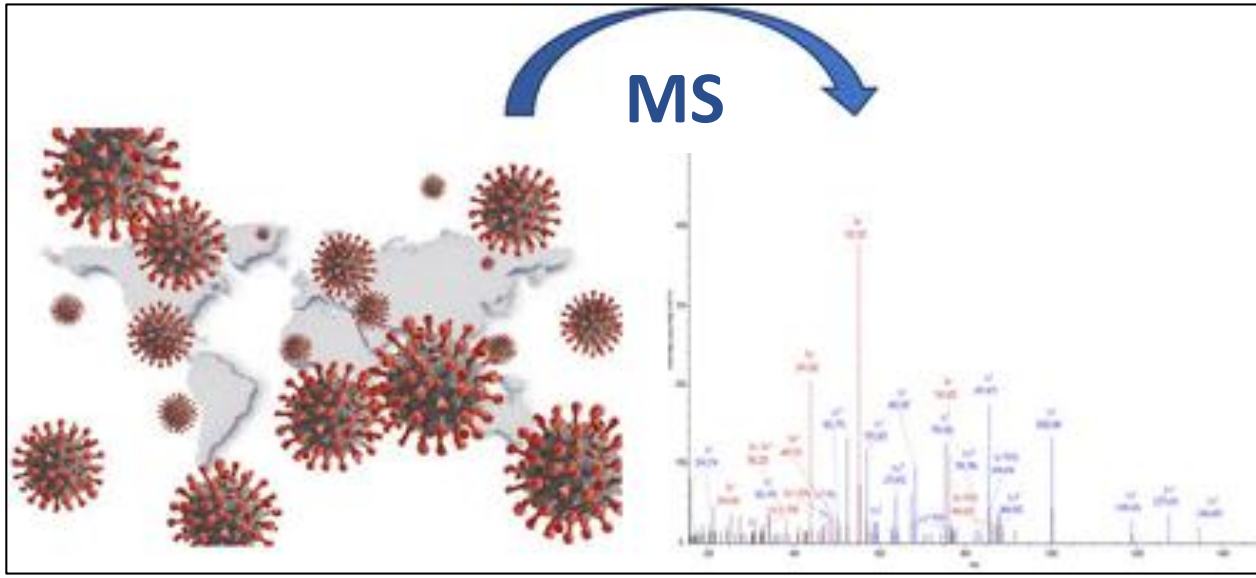



Image credit (left): Gerd Altmann, Pixabay License, https://pixabay.com/illustrations/corona-coronavirus-virus-covid-19-4959447

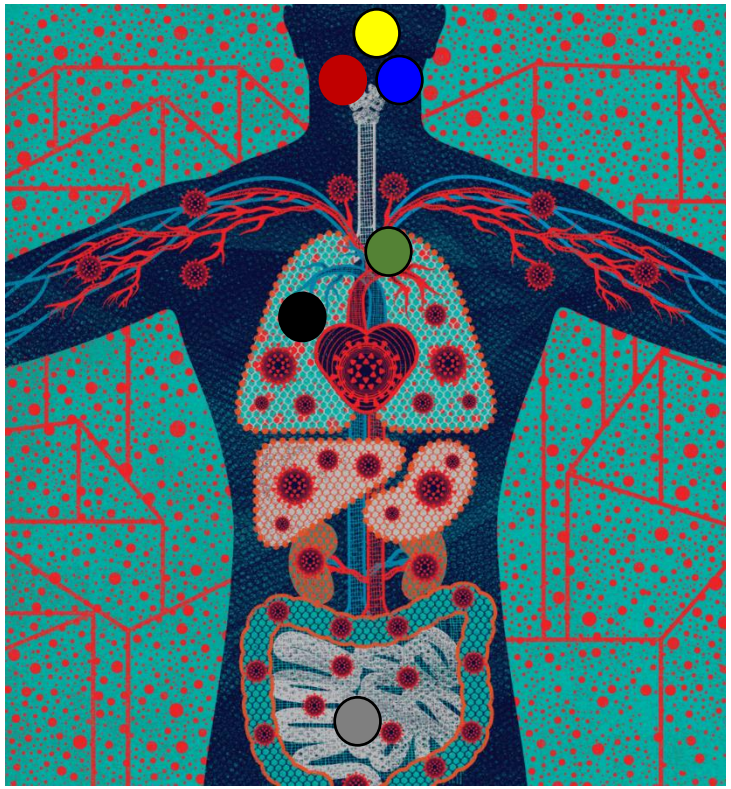
GalaxyP
galaxyp.org

## CLINICAL SAMPLES

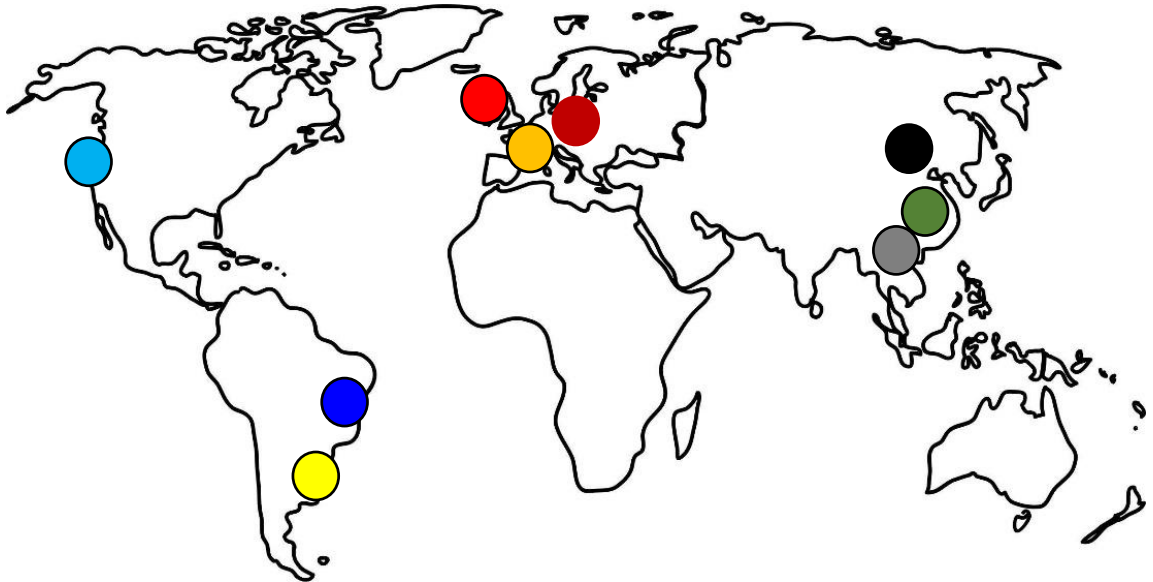
| Dataset | ProteomeXchange ID \| Pubmed ID | Lab |
|---|---|---|
| **Gargling Solution** | PXD019423 \| PMID: 32568543 | Sinz Lab (Halle, Germany) |
| **Nasopharyngeal swabs** | PXD020394 \| PMID: 32835036 | Lima Lab (Montevideo, Uruguay) |
| **Respiratory tract samples** | PXD021328 \| PMID: 33273458 | Carvalho Lab (São Paulo, Brazil) |
| **Broncheo-alveolar lavage fluid (BALF)** | PXD022085 \| PMID: 33098359 | Cheng Lab (Wuhan, China) |
| **Lung Samples** | PXD018094 \| PMID: 33060566 | Zhong Lab (Beijing, China) |
| **Gut Microbiome** | PXD023099 \| Unpublished | Yan Lab (Guangzhou, China) |

## CELL CULTURE

| Dataset | ProteomeXchange ID \| Pubmed ID | Lab |
|---|---|---|
| **Time series** | PXD018594 \| PMID: 32619390 | Armengaud Lab (Bagnols-sur-Cèze, France) |
| **8 hours time point** | PXD018804 \| PMID: 32462744 | Armengaud Lab (Bagnols-sur-Cèze, France) |
| **Proteo-transcriptomics analysis** | PXD018241 \| PMID: 32723359 | Matthews Lab (Bristol, UK) |
| **Host-viral protein interaction** | PXD018117 \| PMID: 32353859 | Krogan Lab (San Franscisco, CA) |



https://www.ucsf.edu/magazine/covid-body

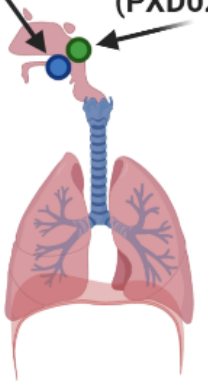


# COVID-19 DETECTION MASS SPECTROMETRY METHODS

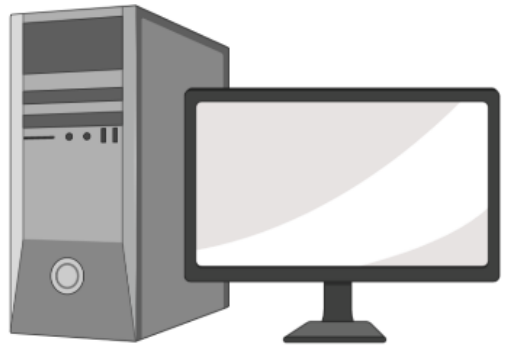
***In Vivo* Datasets**

Saline gargling
(PXD019423)

Oro/nasopharyngeal swabs
(PXD021328)

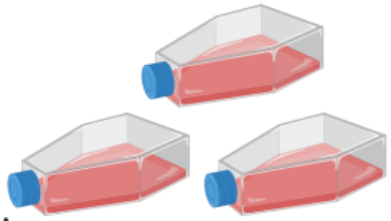


**Bioinformatically-derived peptides**
**(Orsburne *et al.*)**



***In Vitro* Datasets**

Gouveia *et al.* (PXD018804)
Grenga *et al.* (PXD018594)
Davidson *et al.* (PXD018241)



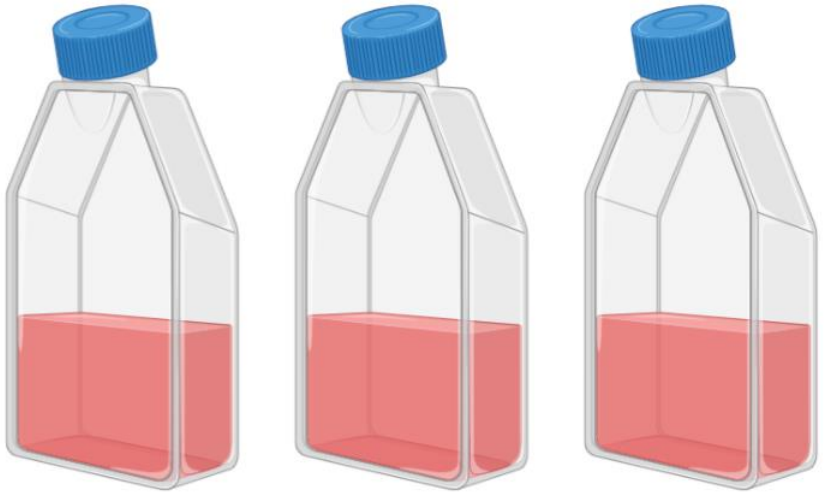
Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines. Grenga *et al*; Emerg Microbes Infect 9(1):1712-1721. doi:10.1080/22221751.2020.1791737. **May 2020**

Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. Ihling *et al* J Proteome Res. 6;19(11): 4389-4392. doi: 10.1021/acs.jproteome.0c00280. **April 2020**

In silico approach toward the identification of unique peptides from viral protein infection: Application to COVID-19. Orsburn *et al* doi: https://doi.org/10.1101/2020.03.08.980383 **April 2020**

# Multiple datasets were used in the creation of a peptide panel and the validation of their utility in diagnosing SARS-CoV-2

PXD018804, PXD018594, PXD018241



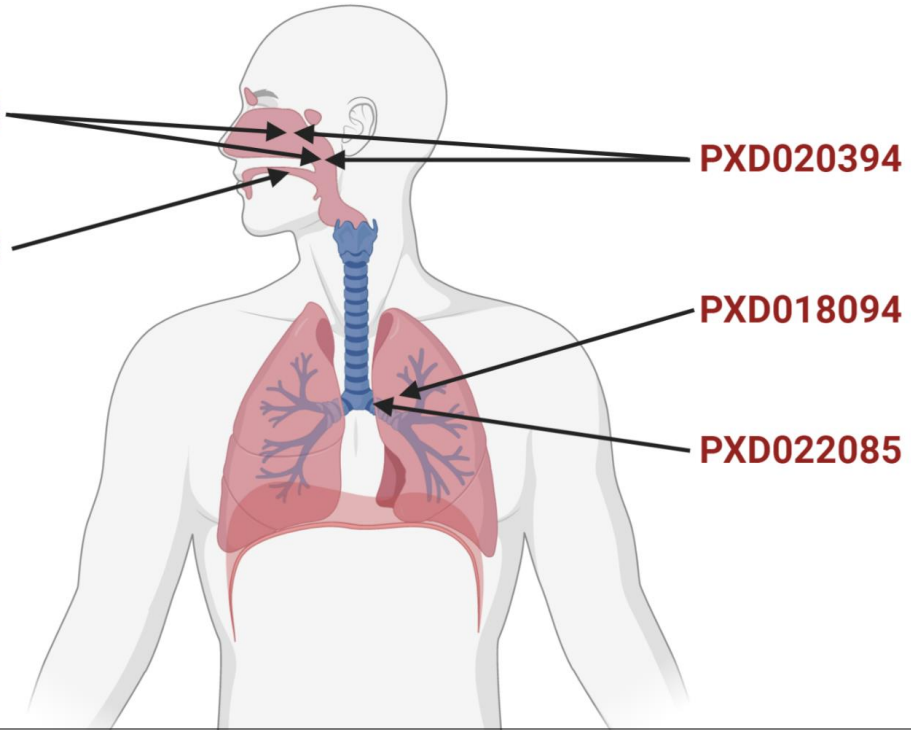
PXD021328

PXD020394

PXD019423

PXD018094

PXD022085



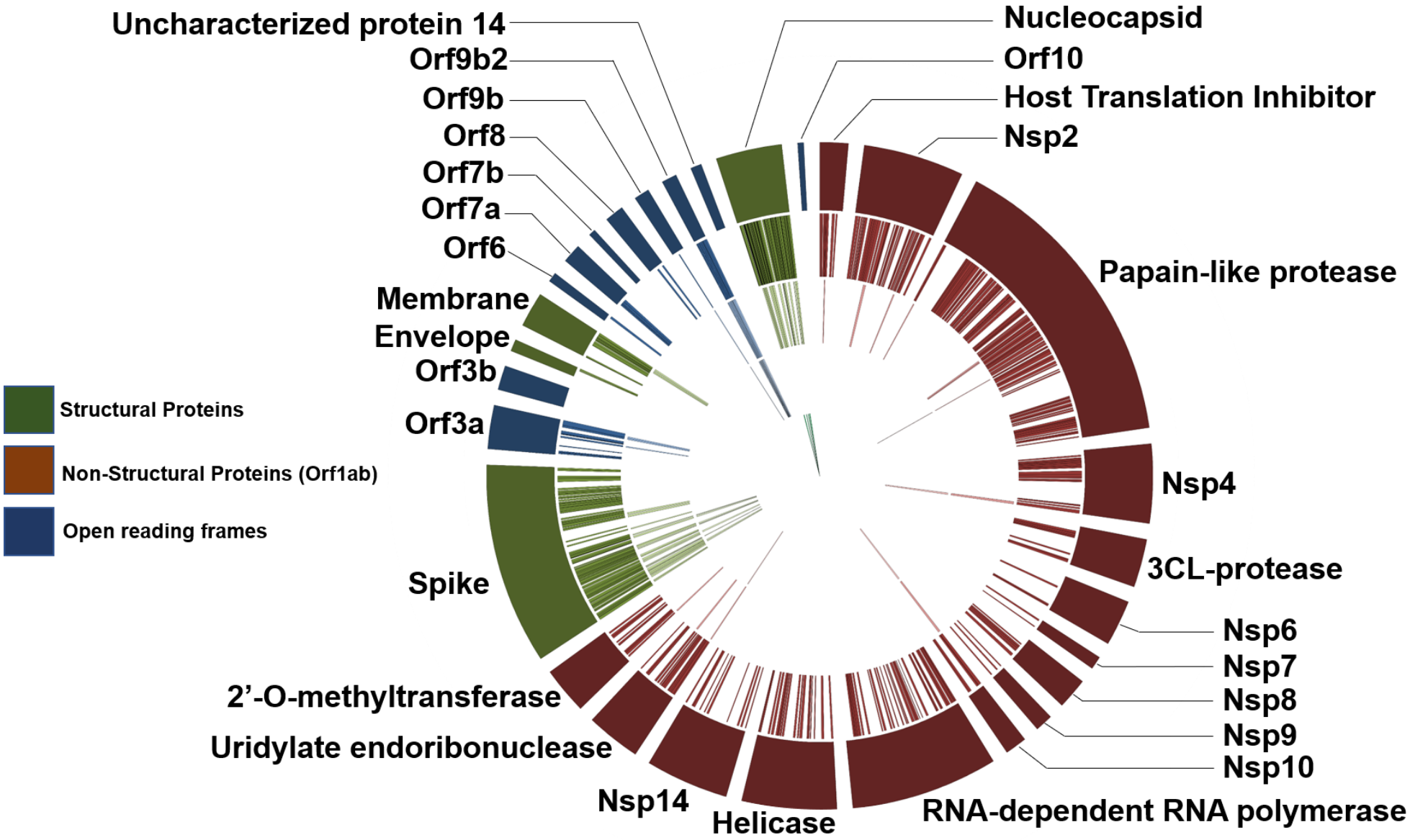
**Peptide Panel Generation**

**Peptide Validation**

## Database Search Workflow

## Peptide Validation Workflow

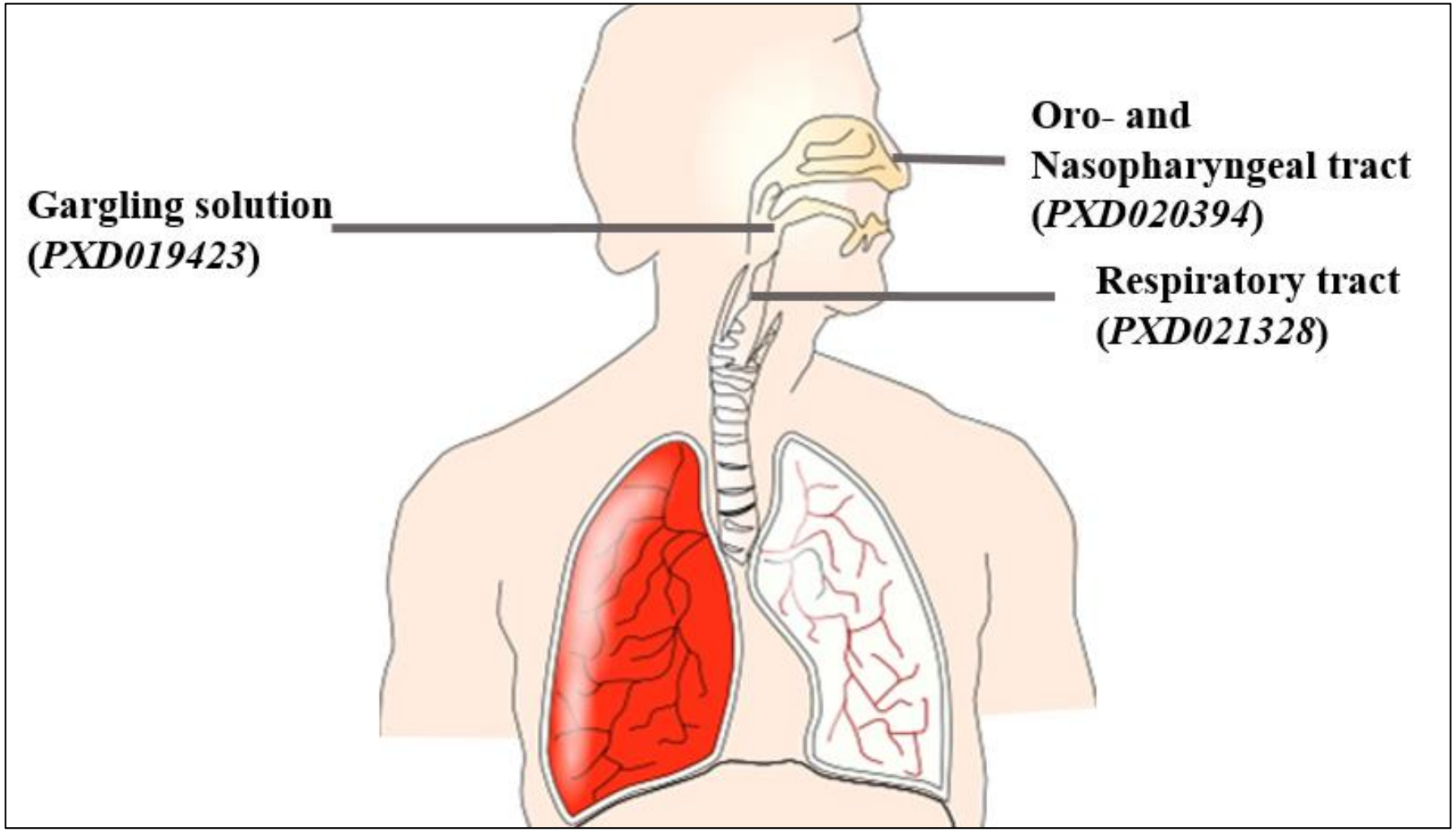
# Protein assignment of detected and validated SARS-CoV-2 peptides




Rajczewski *et al* (2021) Accepted for publication in Clin. Proteomics.
MedRxiv Preprint: https://www.medrxiv.org/content/10.1101/2021.02.09.21251427v2

GalaxyP
galaxyp.org

# Metaproteomics analysis of SARS-CoV-2 infected patient samples




Thuy-Boun *et al (2021) Journal of Proteome Research*
*http://dx.doi.org/10.1021/acs.jproteome.0c00822*

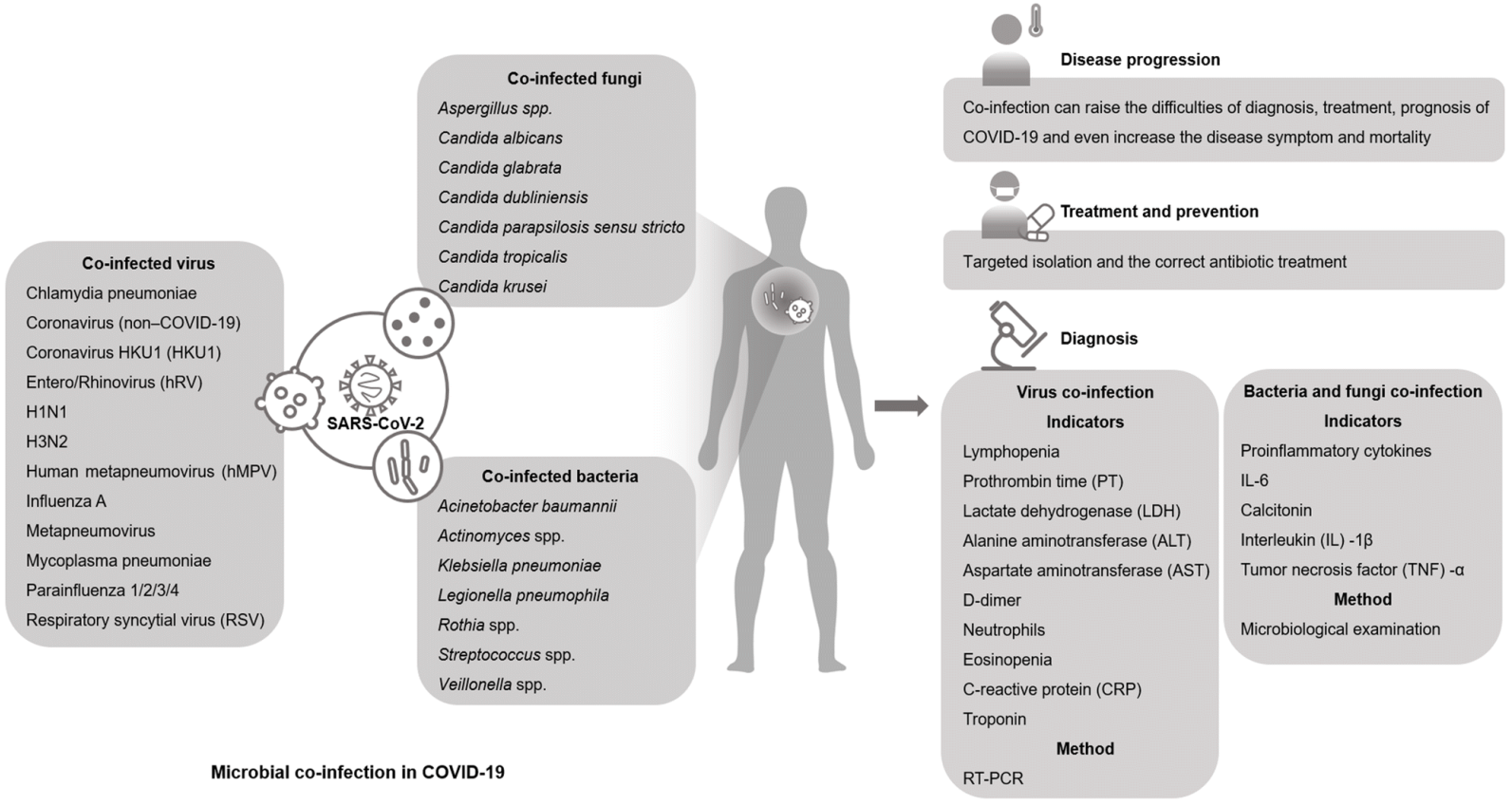
# Co-Infection in COVID-19 Patients

Zhu *et al* **Co-infection with respiratory pathogens among COVID-2019 cases**. Virus Res . (2020) 285:198005.

Chen *et al* **The microbial coinfection in COVID-19** Appl Microbiol Biotechnol . (2020) 104(18):7777-7785.

Mirzai *et al* **Bacterial co-infections with SARS-CoV-2** IUBMB Life . (2020) 72(10):2097-2111.

Bao et al **Oral Microbiome and SARS-CoV-2: Beware of Lung Co-infection**. Front Microbiol . (2020);11:1840.



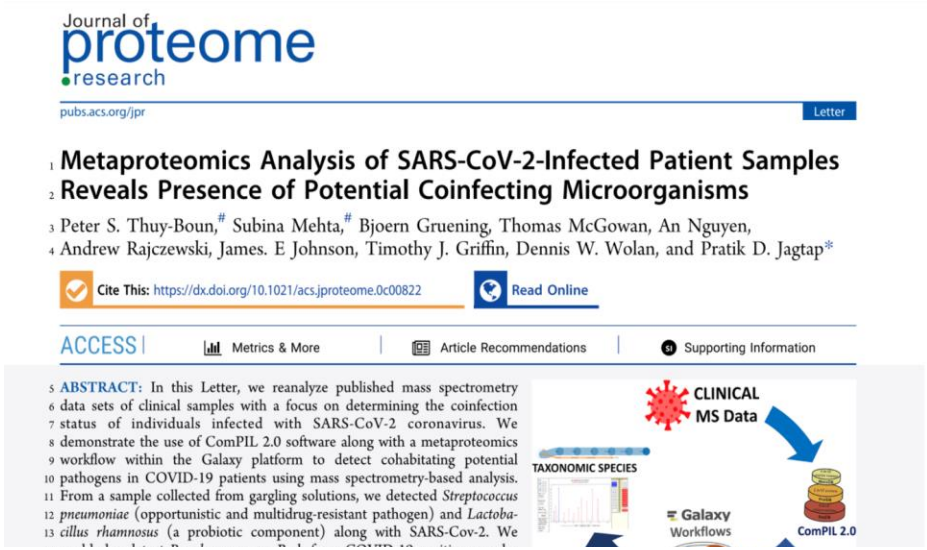

https://link.springer.com/article/10.1007/s00253-020-10814-6

- **Co-infection has an effect on the diagnosis, symptoms, treatment and mortality.**
- **Patient could be infected prior to COVID-19 infection or during hospitalization.**
- **Nosocomial infections can affect antibiotics treatment plans due to antibiotic resistance.**
- **Culture-based detection methods prolong diagnosis of the disease.**

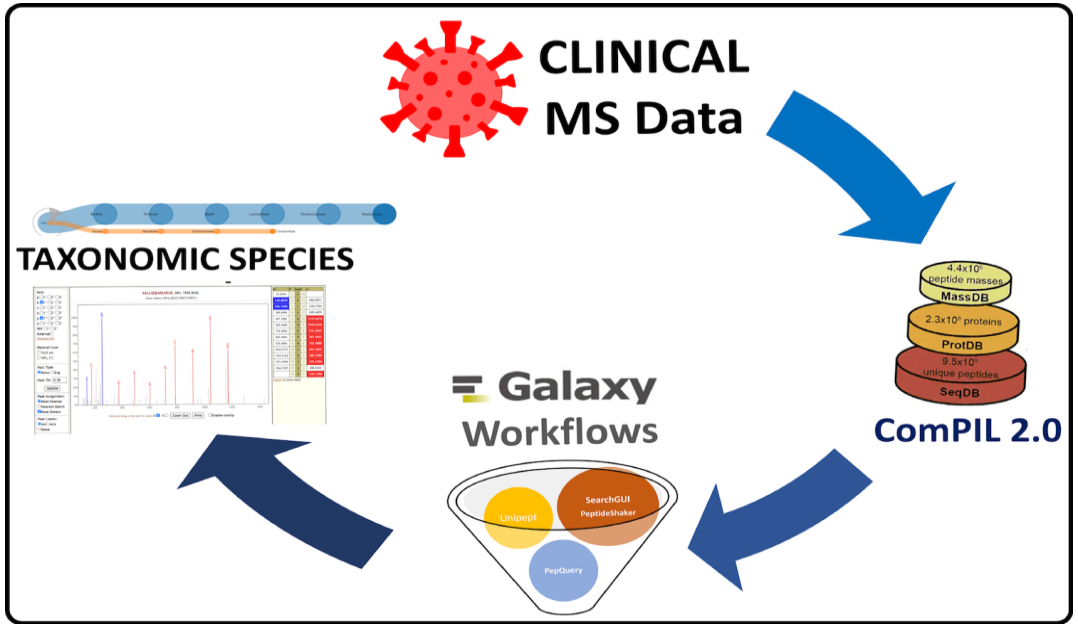
GalaxyP
galaxyp.org

# Datasets and organisms detected

| | |
|---|---|
| ***Streptococcus pneumoniae*** | **Causes pneumonia (respiratory-tract infection)** |
| ***Lactobacillus rhamnosus*** | **Probiotic** |
| ***Pseudomonas sp.* BcH** | **Unclassified *Pseudomonas*** |
| ***Pseudomonas monteilii*** | **Meningoencephalitis** |
| ***Acinetobacter ursingii*** | **Bacteremia** |




Thuy-Boun *et al (2021) Journal of Proteome Research*
*http://dx.doi.org/10.1021/acs.jproteome.0c00822*

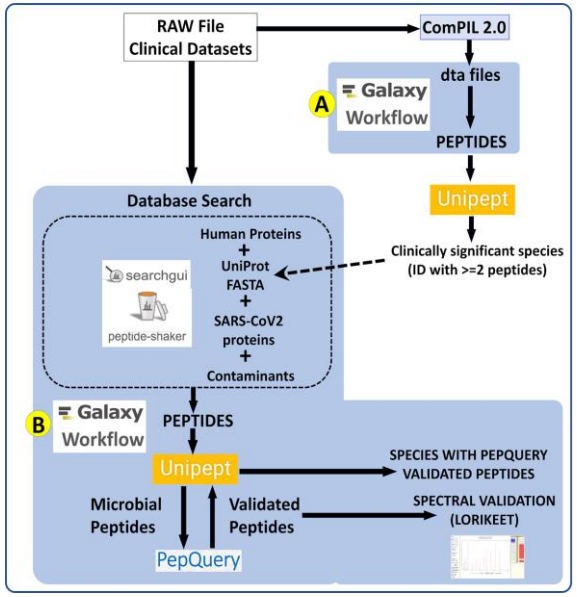


# https://covid19.galaxyproject.org/proteomics/

- Galaxy workflows are available for analysis of COVID-19 MS datasets.
- We could detect peptides that spanned the SARS-CoV-2 proteome
- Metaproteomics analysis revealed presence of potential co-infecting microorganisms in COVID-19 patient samples.

- Reanalysis of PXD018804
- Reanalysis of PXD018682
- Reanalysis of PXD018117
- Reanalysis of PXD018241
- Reanalysis of PXD018594
- Reanalysis of PXD020394
- Reanalysis of PXD021328
- Metaproteomics of mPXD019423
- Metaproteomics of mPXD021328
- Metaproteomics of mPXD020394

**A rigorous evaluation of optimal peptide targets for MS-based clinical diagnostics of Coronavirus Disease 2019 (COVID-19).**
Rajczewski *et al* (Preprint in MedRxiv)
*https://www.medrxiv.org/content/10.1101/2021.02.09.21251427v1*
***ACCEPTED FOR PUBLICATION IN CLINICAL PROTEOMICS***




Peter Thuy-Boun *et al*
*http://dx.doi.org/10.1021/acs.jproteome.0c00822*

GalaxyP
galaxyp.org

# HANDS-ON SESSION

**Subina Mehta**

**Andrew Rajczewski**

**Dinh Duy An Nguyen**

## Instructions

Please **Register** for creating an account with a valid email ID and Password at [usegalaxy.eu](usegalaxy.eu).

Once Registered, click on [https://usegalaxy.eu/join-training/glbiotraining](https://usegalaxy.eu/join-training/glbiotraining) to join the GLBIO 2021 Galaxy session.

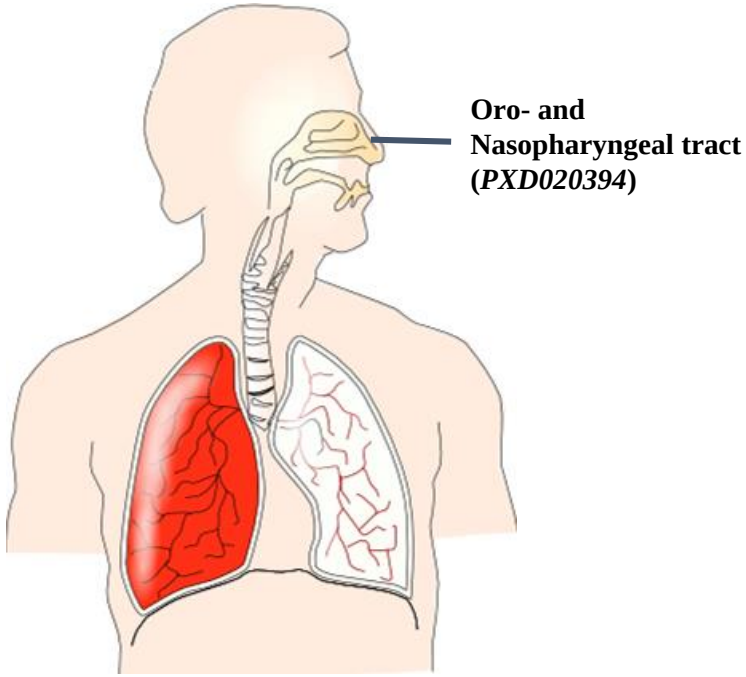
Go to Shared Data
➡ Histories
➡GLBIO inputs
Import +

Go to Shared Data
➡ Workflows.
➡GLBIO workflow
Import +

Run the workflow on active history

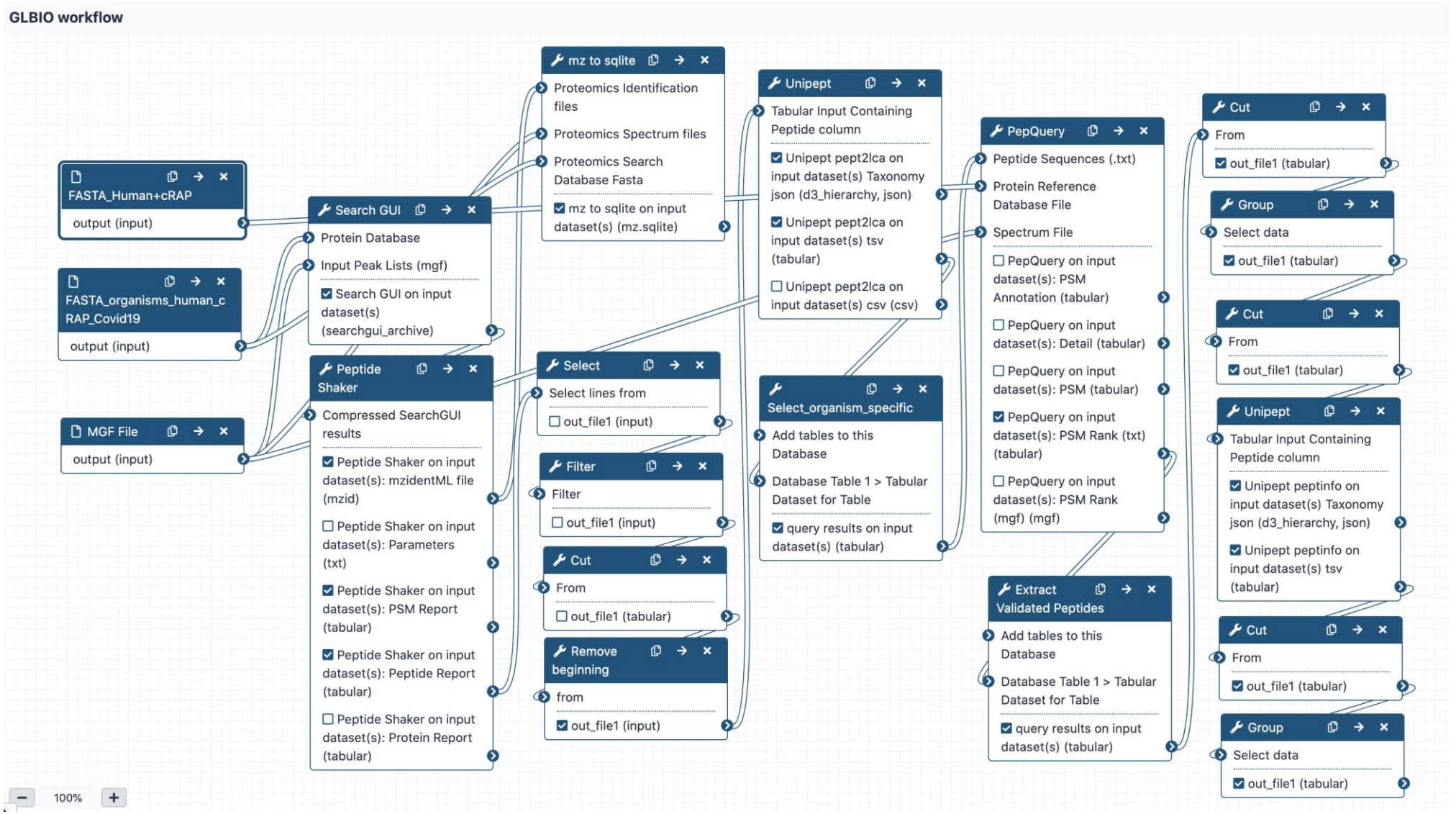
# Determining the presence of COVID-19 peptides in Galaxy

# Dataset Information



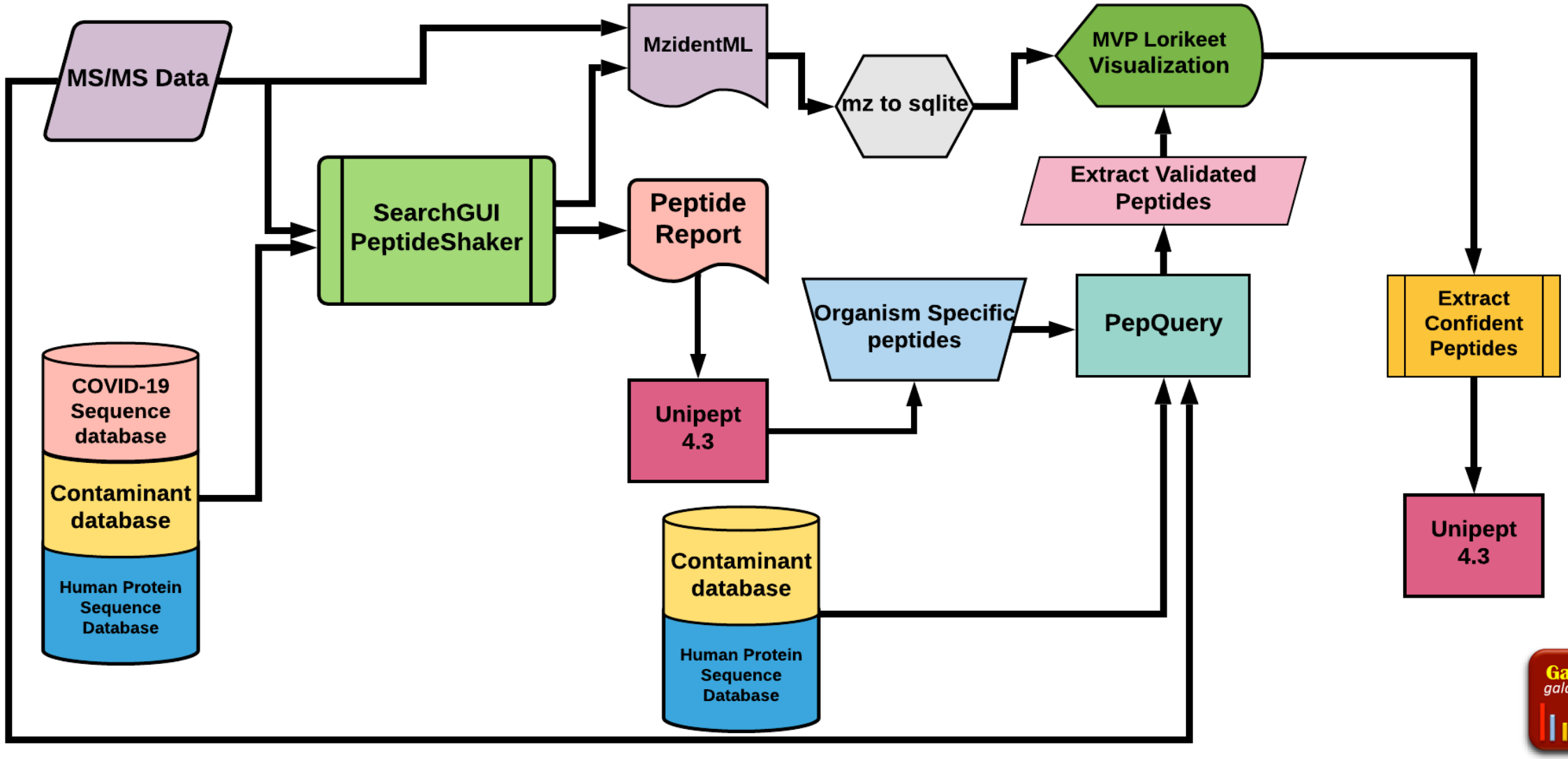

- Oro-Nasopharyngeal
- Patient data - Positive and Negative
- Coronavirus present?
- Any other species – co-infection?

# GALAXY WORKFLOW




GalaxyP
galaxyp.org

# Analytical Workflow



# Instructions

- Go to Shared Data ➡ Histories ➡ GLBIO workshop Sample 1 ➡ Import ➕

OR

- Go to Shared Data ➡ Histories ➡ GLBIO workshop Sample 2 ➡ Import ➕

# SearchGUI-Peptide Shaker

1%FDR

TABULAR OUTPUTS

PSM REPORT, PEPTIDE REPORT, PROTEIN REPORT

MZIDENTML FILE

# Unipept

Match taxa

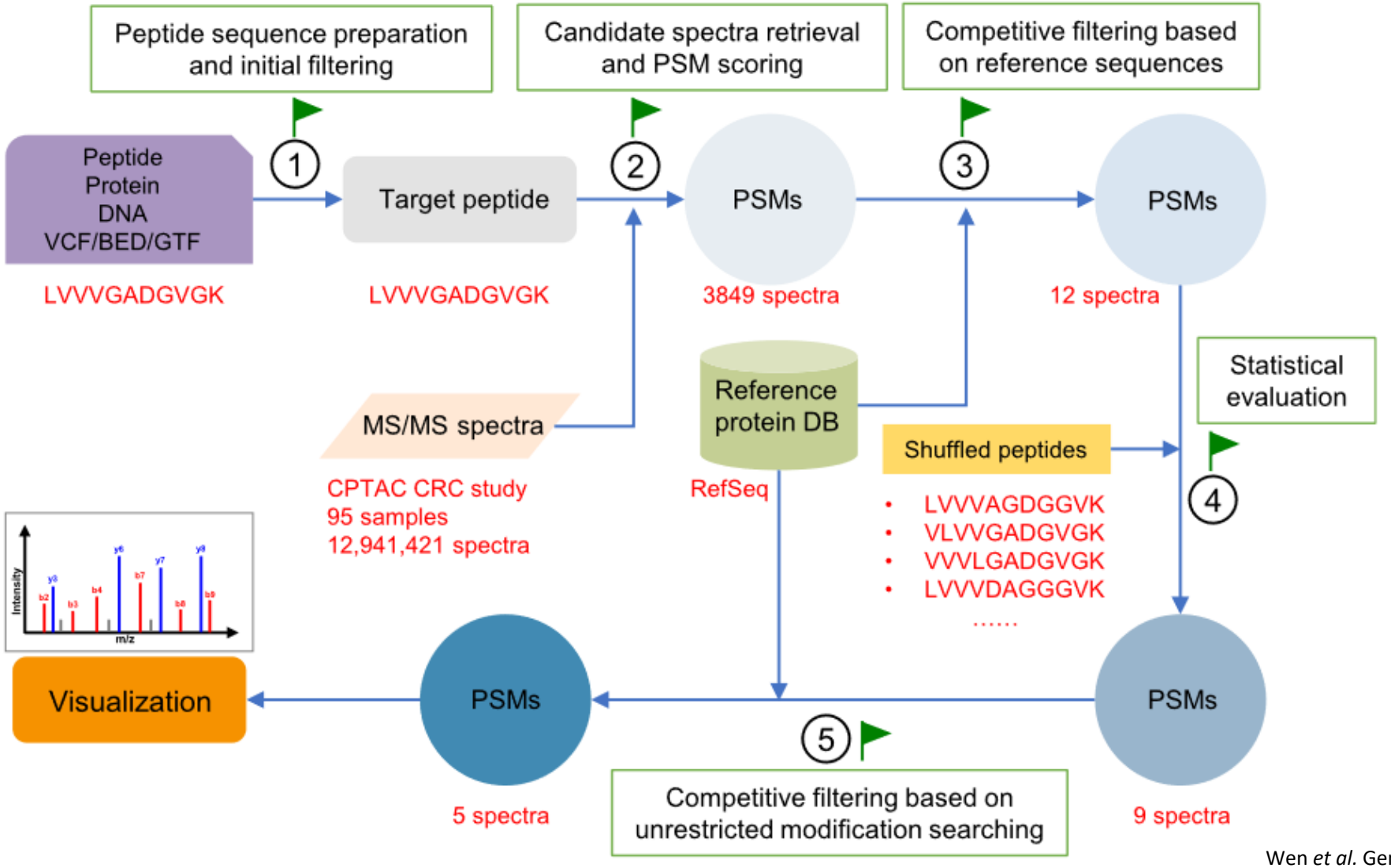
select organisms with at least 2 peptides

# Peptide Validation

PepQuery

Spectral Quality

# PepQuery Workflow




Wen *et al.* Genome Res. 2019. 29: 485-493

# PepQuery

Features

PSM rank

Eliminate peptides in reference database

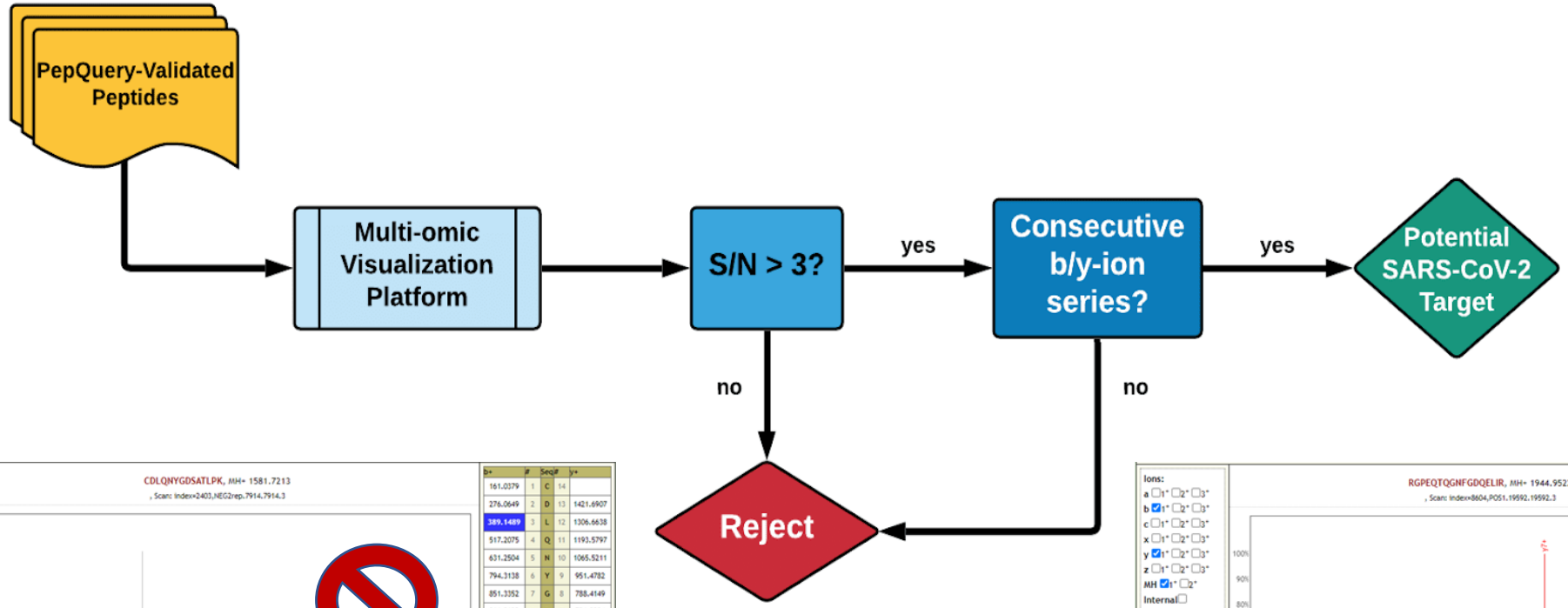
P-value

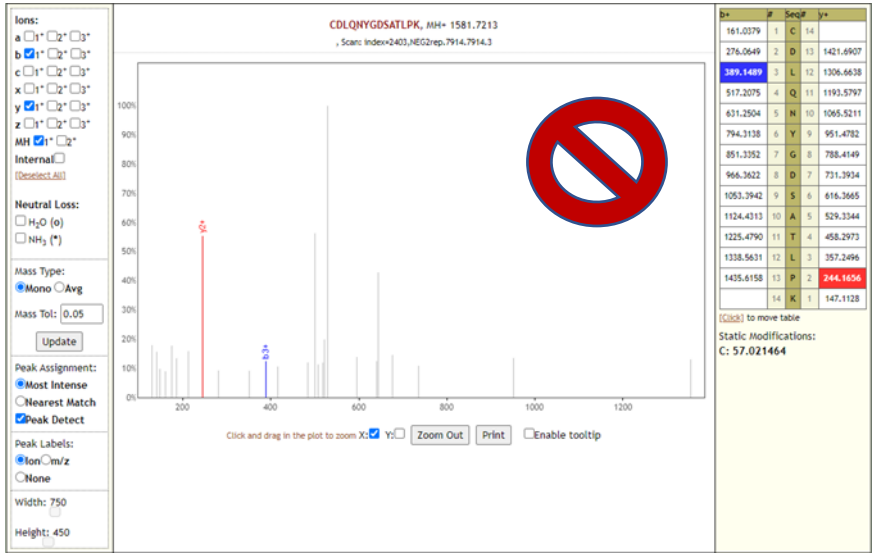
confidence

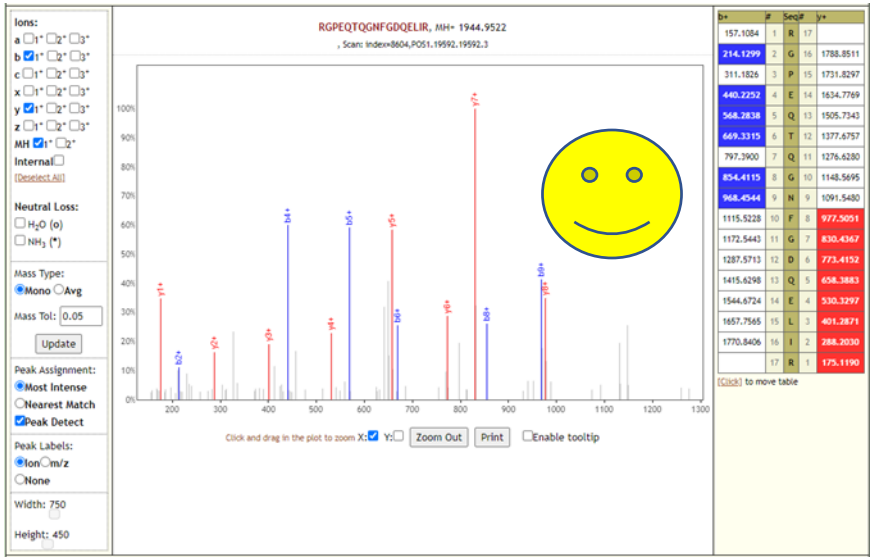
# Discussion Point!

What do you think is the COVID and co-infection status of this sample?

# Spectral Validation (Lorikeet)



# Thank you!

*Questions?*

***Useful links:***

- *galaxyp.org*
- *proteomics.usegalaxy.eu*
- *Galaxy Workflows for COVID-19 Research*
- *Galaxy Training Network*
- *Co-infection Metaproteomics*
- *Peptides for clinical diagnosis of COVID-19*