

# **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](http://galaxyp.org)  
[z.umn.edu/itcrgalaxyvideo](http://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

# 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**



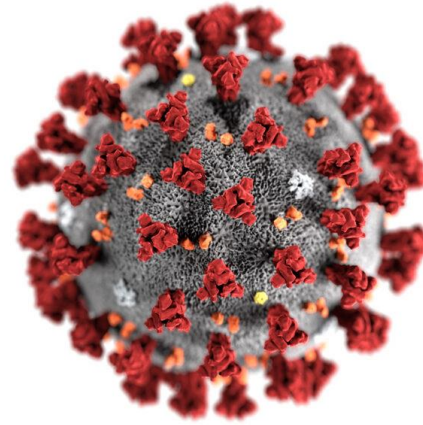
**NATIONAL CANCER INSTITUTE**  
Informatics Technology for  
Cancer Research



# Topic(s) of workshop



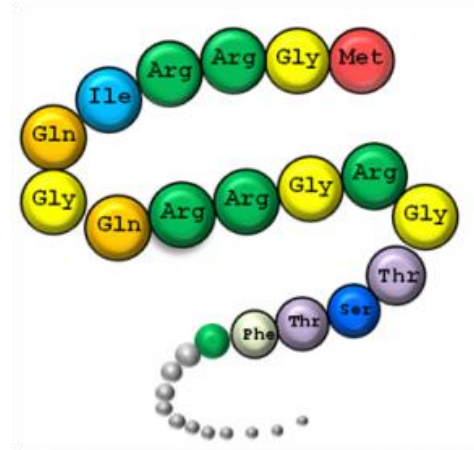
*Mass spectrometry-based proteomic informatics*



*COVID-19 research*

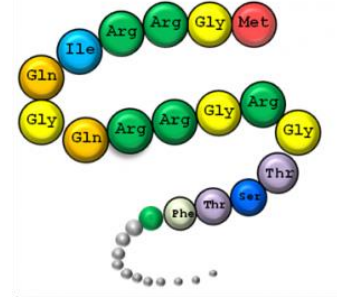


*Metaproteomics (multi-omics)*



*Protein sequence analysis*

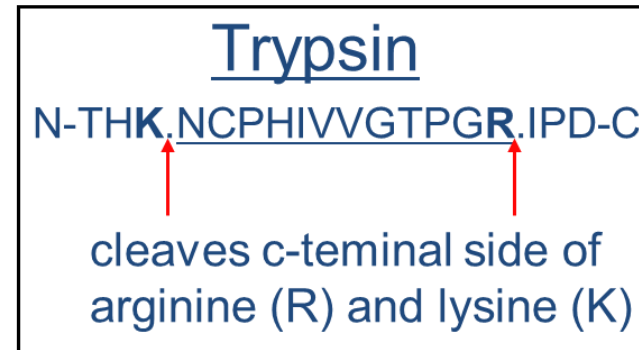
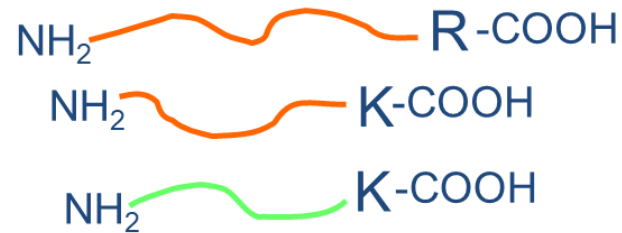
# Shotgun proteomics: breaking proteins into peptides



1. Protein, MW = 10,000 + 

digest into peptides (enzymatic, chemical)

2. Peptides,  
MW < 4,000

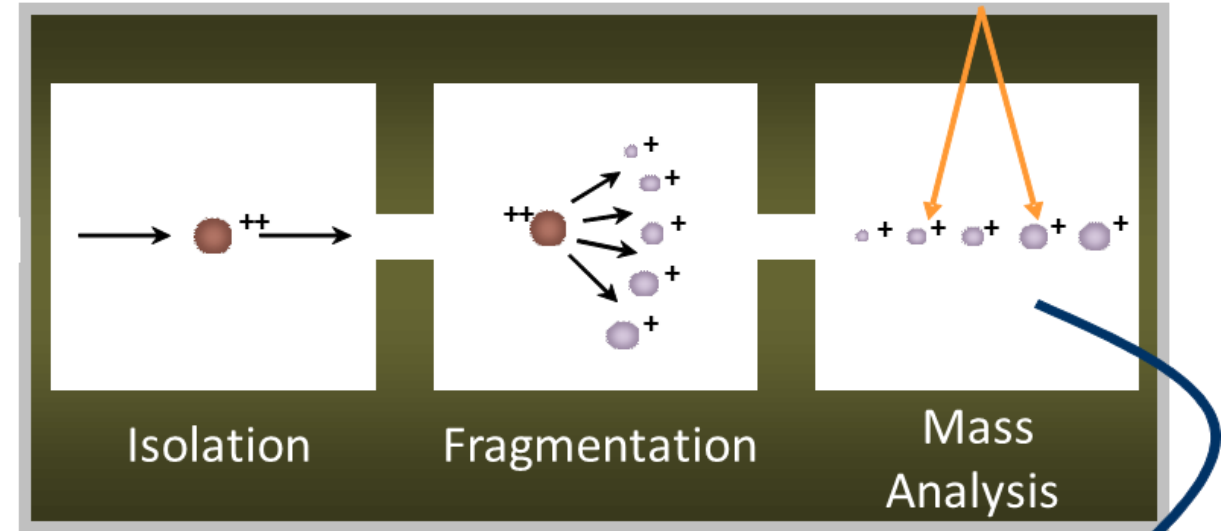
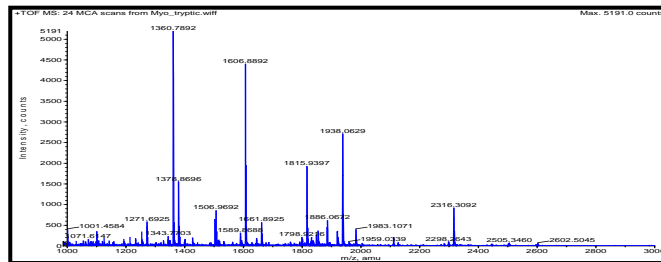
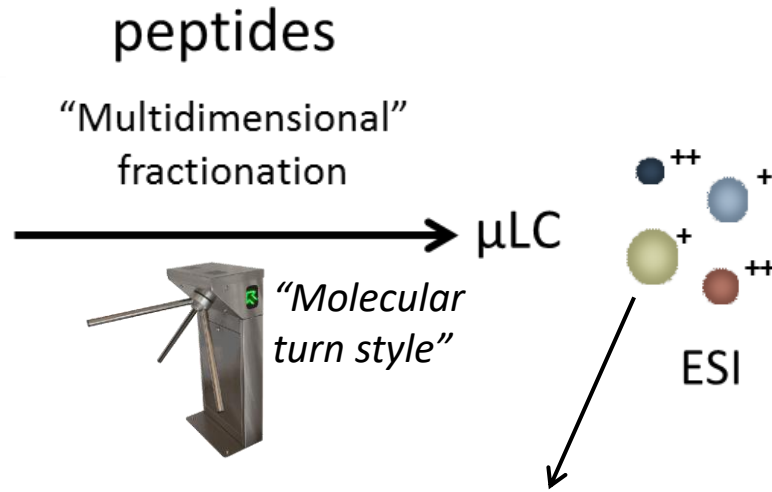
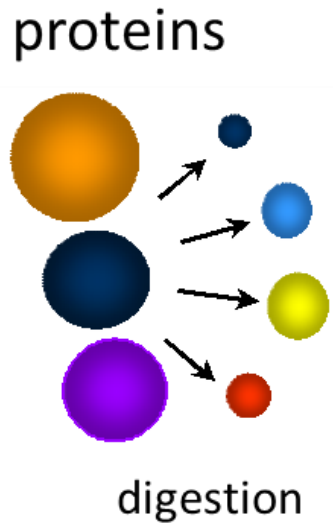


3. Peptide  
fragments

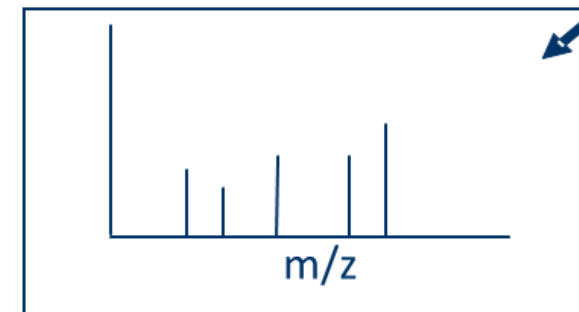


# Mass Spectrometry-based proteomics: A primer

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

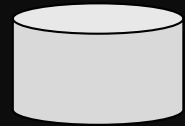
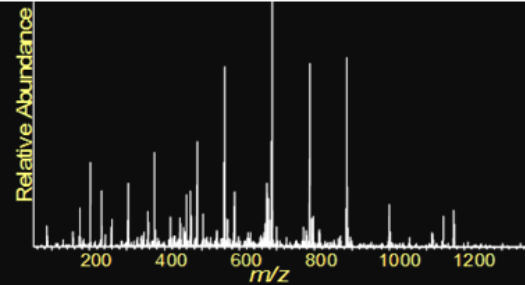


MS2



# Matching amino acid sequences to MS/MS data

Raw MS/MS spectrum

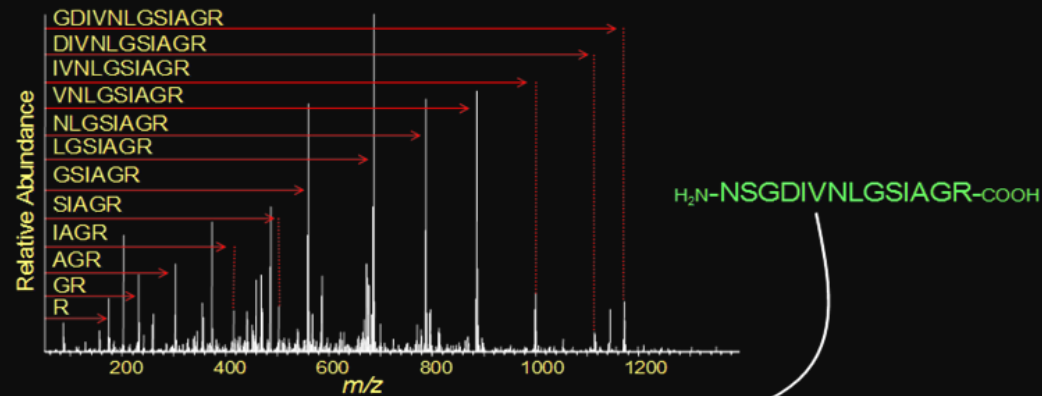


Protein sequence and/or DNA  
sequence database search



Direct identification of 1000s  
proteins from complex mixtures

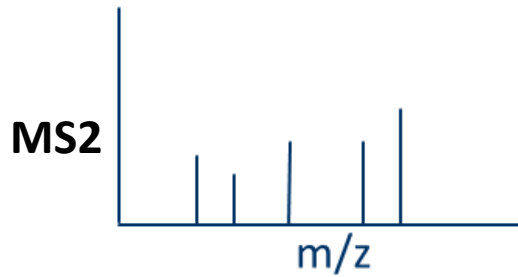
Peptide sequence match



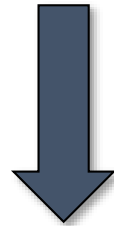
Protein identification



# Inferring protein identity for peptide sequences



CAQCHTVEK



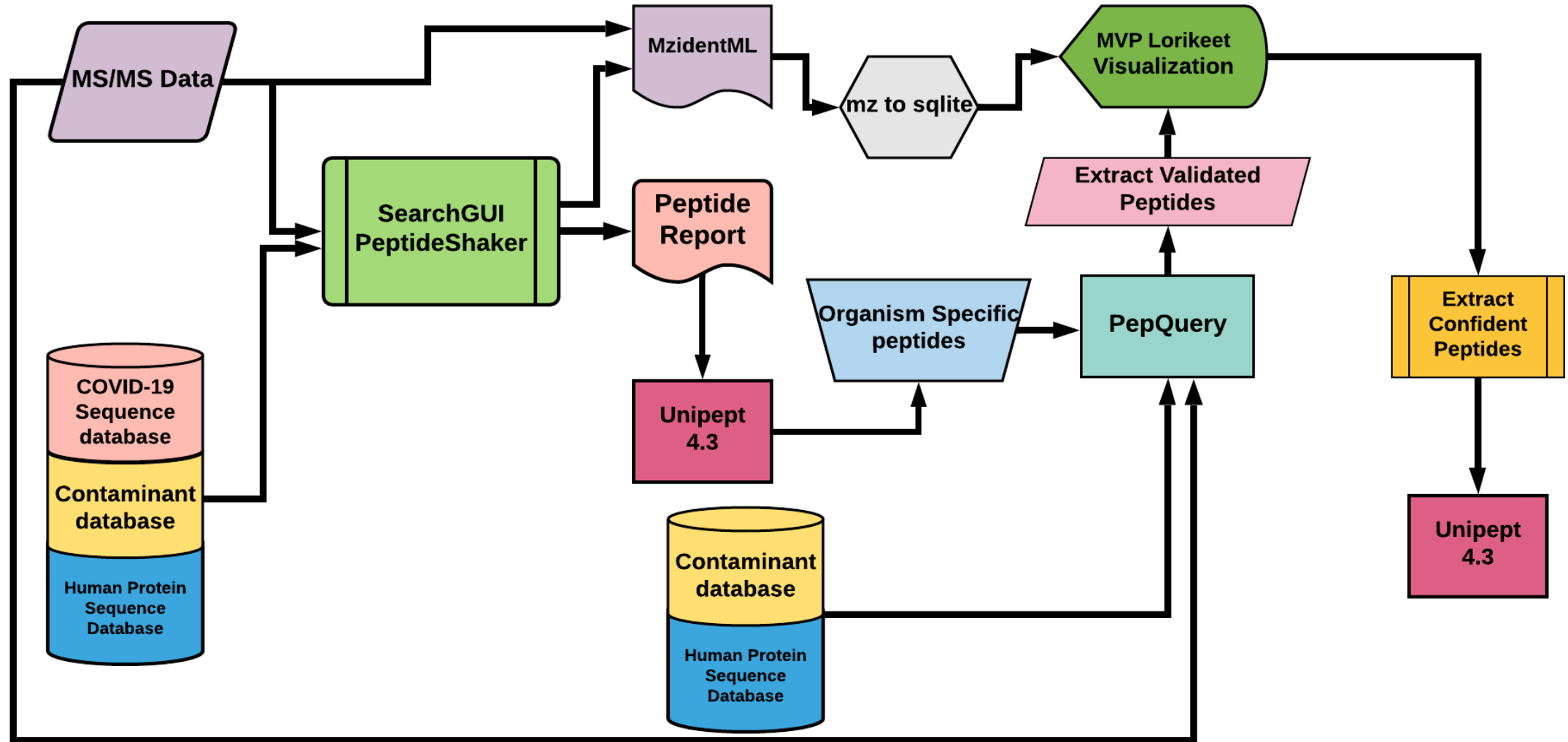
**Cytochrome C**

NH<sub>2</sub>GDVEKGGKIFVQK**CAQCHTVEK**GGKHKTGPNLHGLFG**RKTGQAPGFTYTDANKNKGITW**  
**KEETLMEYLENPK**KYIPGTKMIFAGIKKKTER**EDLIAYLK**KATNE<sub>COOH</sub>



# Driven by bioinformatics

- Many step workflow for analyzing large and diverse datasets



# 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)



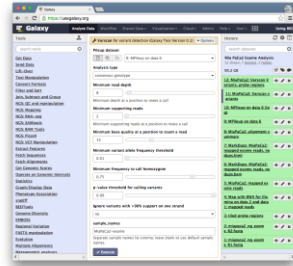
## ***Working philosophy:***



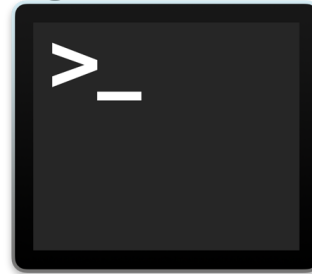
# Galaxy: an integrative workbench well-suited for multi-omics

## Interfaces

Web UI

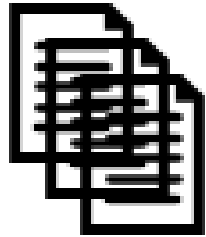


Programmatic API



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

## Datasets



## Computing Resources



## Analysis Tools and Visualizations



*Courtesy Jeremy Goecks, OHSU*

# Accessing Galaxy via a public gateway

- 1 Login/Register: [usegalaxy.eu](https://usegalaxy.eu)
- 2 Go to TlaaS link: <https://usegalaxy.eu/join-training/globiotraining>
- 3 Return to [usegalaxy.eu](https://usegalaxy.eu) site

The screenshot displays the Galaxy Europe web interface. At the top, there is a navigation bar with 'Galaxy Europe' on the left and 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Help', 'Login or Register', and a user icon on the right. Below the navigation bar, the main content area is divided into several sections:

- Tools:** A sidebar on the left contains a search bar, an 'Upload Data' button, and a list of tool categories including 'GENERAL TEXT TOOLS', 'GENOMIC FILE MANIPULATION', 'FASTA/FASTQ', 'FASTQ Quality Control', 'Quality Control', 'SAM/BAM', 'BED', 'VCF/BCF', 'Nanopore', 'COMMON GENOMICS TOOLS', 'GENOMICS ANALYSIS', and 'Assembly'.
- COVID-19 Research!:** A green banner with a 'Log in or register a new account' button. The text encourages users to learn best practices for SARS-CoV-2 data analysis and mentions a dedicated training portal.
- News:** A section with several news items, including 'May Galactic News!', 'GCC2021 Abstract and Fellowship Submission, and Early Registration Deadlines Extended', 'Plant Transcriptomics Workshop: A Summary', 'Monitoring of raw open data for SARS-CoV-2 genome surveillance', and 'UseGalaxy.eu Tool Updates for 2021-04-26'.
- Events:** A section with several event items, including 'Galaxy Resources for Tool Developers', 'Customizeable welcome and introduction page for new Galaxy users', 'Galaxy Training Network day - (online) CoFest and community call', 'Galaxy Paper Cuts', 'Metatranscriptomics analysis using microbiome RNA-seq data in Galaxy', and '9th RepeatExplorer Workshop on the Application of Next Generation Sequencing to Repetitive DNA Analysis'.
- History:** A sidebar on the right showing 'search datasets' and 'Unnamed history (empty)'. A blue box indicates that the history is empty and suggests loading data from an external source.

At the bottom of the page, there are three numbered troubleshooting tips in orange text:

1. This could be caused by your reverse proxy settings.
2. If you host grafana under subpath make sure your grafana.ini root\_url setting includes subpath. If not using a reverse proxy make sure to set serve\_from\_sub\_path to true.
3. If you have a local dev build make sure you build frontend using: yarn start, yarn start:hot, or yarn build

An 'OPEN CHAT' button is located at the bottom right of the page.



# The Galaxy Interface

Main viewing window  
(workflow development, results visualization etc)

Tools

The screenshot displays the GalaxyP web interface. At the top, a navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The main content area features a green 'Welcome to GalaxyP' message with a checkmark icon, explaining that GalaxyP is a multiple 'omics' data analysis platform. Below this is an 'Updates' section with a 'GalaxyP' logo and a bar chart, listing updates from February 13, 2015, and February 12, 2015, both stating 'All Galaxy Tools running nominally'. On the left, a 'Tools' sidebar lists various tool categories such as 'CORE TOOLS', 'PROTEOMICS', and 'Visualizers'. On the right, a 'History' panel shows a list of workflow steps, including 'imported: Test1 Workflow for human proteogenomics' and several 'BLAST-P Filtered Peptide Report' and 'Regex Find And Replace' steps.


History



# Built for the community: Tool sharing and access



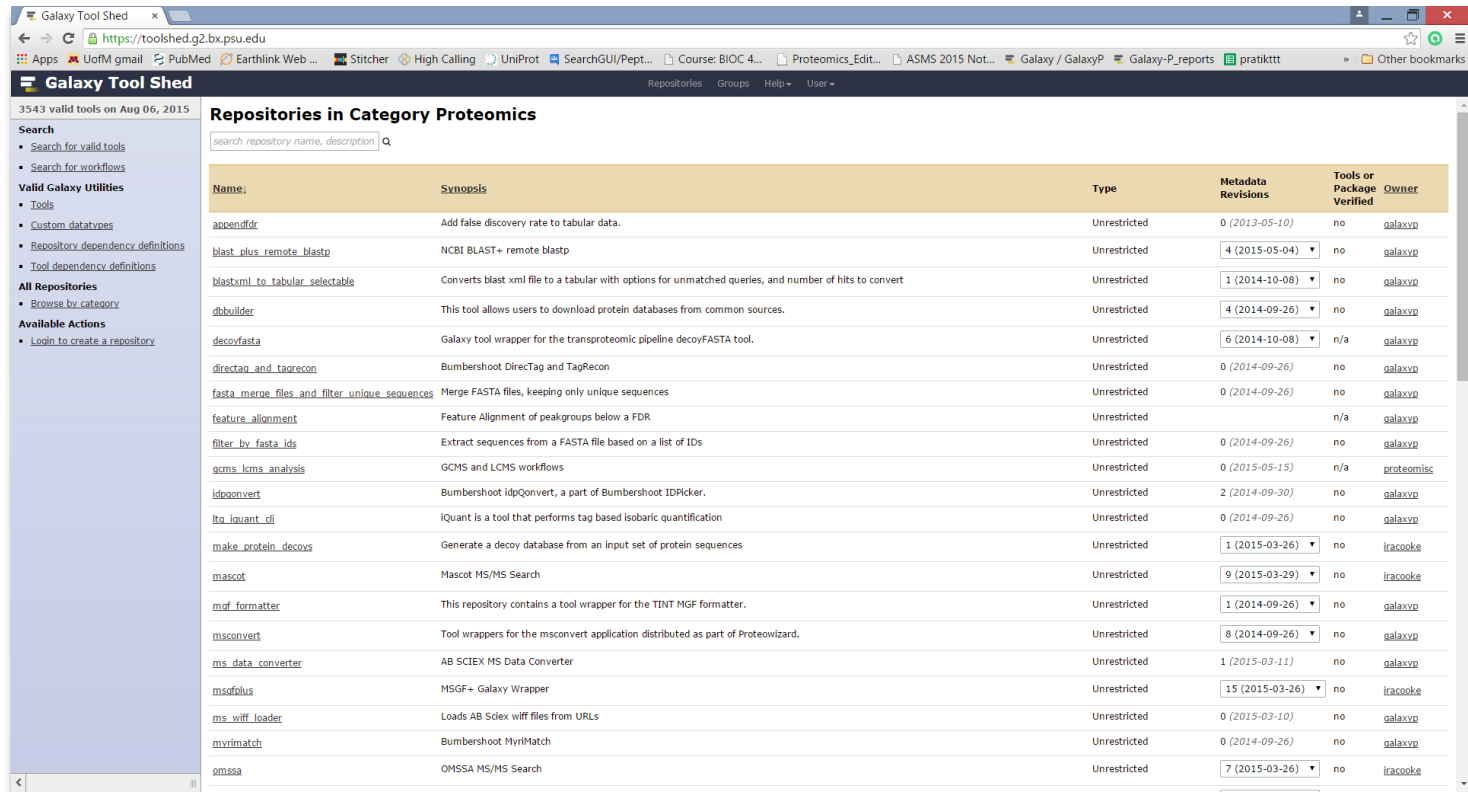
Blankenberg et al. *Genome Biology* 2014, 15:403  
http://genomebiology.com/2014/15/2/403



**OPEN LETTER**

## Dissemination of scientific software with Galaxy ToolShed

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



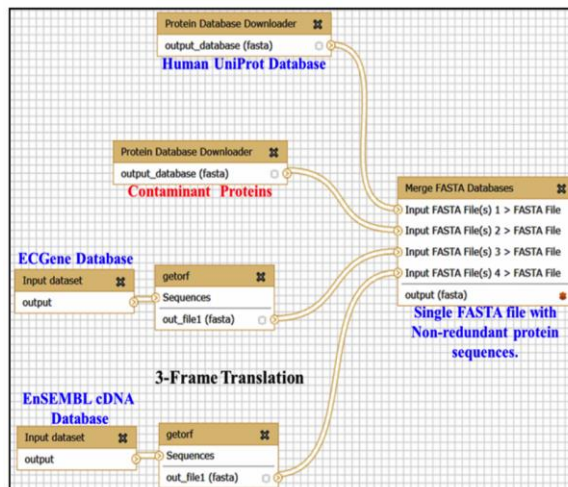
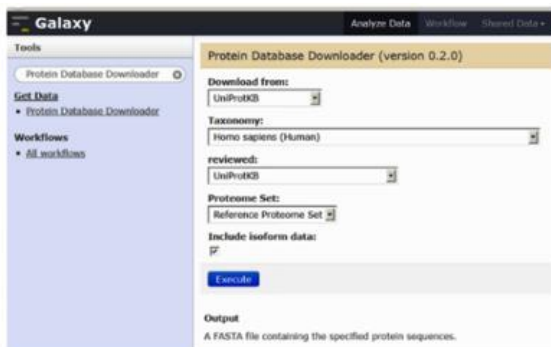
The screenshot shows the Galaxy Tool Shed web interface. The main content area displays a table of repositories in the Proteomics category. The table has columns for Name, Synopsis, Type, Metadata Revisions, Tools or Package Verified, and Owner. The left sidebar contains navigation links for search, utilities, and repositories.

Name:	Synopsis	Type	Metadata Revisions	Tools or Package Verified	Owner
appendfdr	Add false discovery rate to tabular data.	Unrestricted	0 (2013-05-10)	no	galaxy
blast_plus_remote_blastp	NCBI BLAST+ remote blastp	Unrestricted	4 (2015-05-04)	no	galaxy
blastxml_to_tabular_selectable	Converts blast xml file to a tabular with options for unmatched queries, and number of hits to convert	Unrestricted	1 (2014-10-08)	no	galaxy
dbbuilder	This tool allows users to download protein databases from common sources.	Unrestricted	4 (2014-09-26)	no	galaxy
decofasta	Galaxy tool wrapper for the transproteomic pipeline decoyFASTA tool.	Unrestricted	6 (2014-10-08)	n/a	galaxy
directag_and_tagrecon	Bumbershoot DirectTag and TagRecon	Unrestricted	0 (2014-09-26)	no	galaxy
fasta_merge_files_and_filter_unique_sequences	Merge FASTA files, keeping only unique sequences	Unrestricted	0 (2014-09-26)	no	galaxy
feature_alignment	Feature Alignment of peakgroups below a FDR	Unrestricted		n/a	galaxy
filter_by_fasta_ids	Extract sequences from a FASTA file based on a list of IDs	Unrestricted	0 (2014-09-26)	no	galaxy
gcms_lcms_analysis	GCMS and LCMS workflows	Unrestricted	0 (2015-05-15)	n/a	proteomisc
ldpconvert	Bumbershoot ldpConvert, a part of Bumbershoot IDPicker.	Unrestricted	2 (2014-09-30)	no	galaxy
lqant_cli	lQuant is a tool that performs tag based isobaric quantification	Unrestricted	0 (2014-09-26)	no	galaxy
make_protein_decays	Generate a decoy database from an input set of protein sequences	Unrestricted	1 (2015-03-26)	no	iracooke
mascot	Mascot MS/MS Search	Unrestricted	9 (2015-03-29)	no	iracooke
mgf_formatter	This repository contains a tool wrapper for the TINT MGF formatter.	Unrestricted	1 (2014-09-26)	no	galaxy
msonvert	Tool wrappers for the msonvert application distributed as part of Proteowizard.	Unrestricted	8 (2014-09-26)	no	galaxy
ms_data_converter	AB SCIEX MS Data Converter	Unrestricted	1 (2015-03-11)	no	galaxy
msdfplus	MSGF+ Galaxy Wrapper	Unrestricted	15 (2015-03-26)	no	iracooke
ms_wiff_loader	Loads AB Sciex wiff files from URLs	Unrestricted	0 (2015-03-10)	no	galaxy
myrimatch	Bumbershoot MyriMatch	Unrestricted	0 (2014-09-26)	no	galaxy
omssa	OMSSA MS/MS Search	Unrestricted	7 (2015-03-26)	no	iracooke



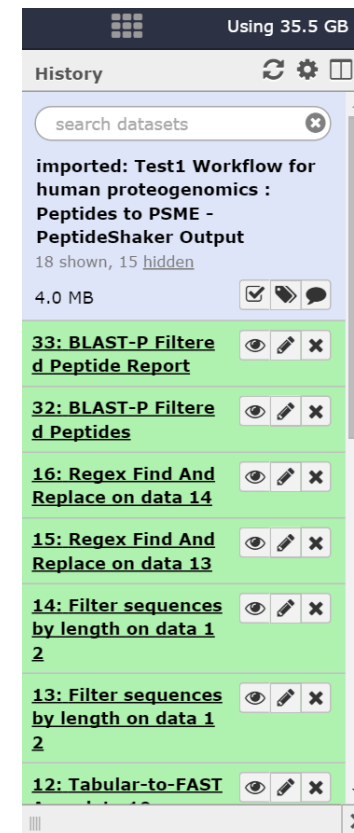
# Defining Workflows and Histories

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

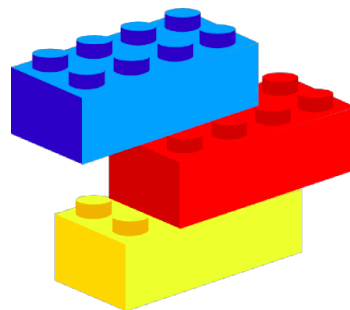


Integrated workflow using multiple, connected tools

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

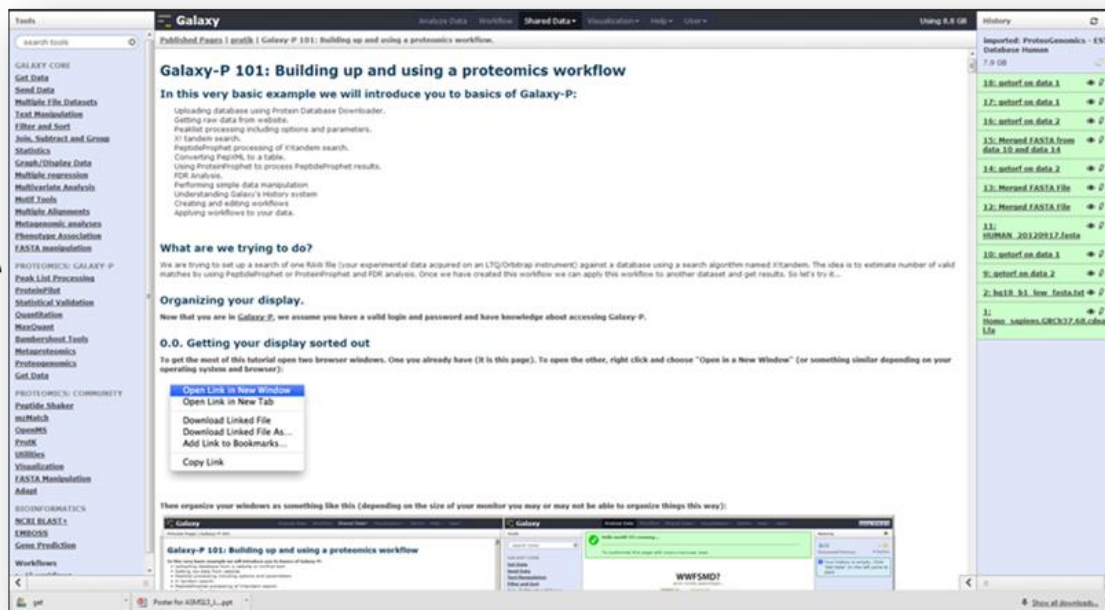


Single software tool in Galaxy





# Built for the community: Sharing workflows and histories



URL  
export

**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>

File export  
(.ga)

Galaxy-Workflow-Workflow\_for\_4-plex\_iTRAQ\_X\_tandem\_Search\_Scaffold\_Processing.ga



# Accessing and using Galaxy



COVID-19 analysis on usegalaxy.★

Genomics Cheminformatics Proteomics Data Webinars Edit on GitHub

## Global platform for the analysis of SARS-CoV-2 data: Genomics, Cheminformatics, and Proteomics

Using open source tools and public cyberinfrastructure for transparent, reproducible analyses of viral datasets.

Powered by: [usegalaxy.org](#) [usegalaxy.eu](#) [usegalaxy.be](#) [usegalaxy.kr](#) [usegalaxy.jp](#)

The goal of this resource is to provide publicly accessible infrastructure and workflows for SARS-CoV-2 data analyses. We currently feature three different types of analyses: **Genomics**, **Cheminformatics**, and **Proteomics**.

### SARS-CoV-2 Monitoring

Why? Workflows Data How To?

### Genomics

Galaxy / Proteomics

Analyze Data Workflow Visualize Shared Data Help Login or Register Using 0 bytes

### Tools

search tools

Upload Data


- Get Data
- Send Data
- Collection Operations
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- GENOMIC FILE MANIPULATION
- Convert Formats
- FASTA/FASTQ
- FASTQ Quality Control
- Quality Control
- SAM/BAM
- BED
- VCF/BCF
- Nanopore
- COMMON GENOMICS TOOLS
- Operate on Genomic Intervals
- Fetch Sequences / Alignments
- GENOMICS ANALYSIS
- Annotation
- Multiple Alignments

#### COVID-19 Research!

Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror all public SARS-CoV-2 data from ENA in a Galaxy data library for your convenience. The Galaxy community has created COVID-19 dedicated training materials. Please check our recent activities for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.

### Welcome to Galaxy Proteomics



Visit [galaxy.org](#) for more info

This European Galaxy branches off from the main [usegalaxy.eu](#) server and focuses on one specific domain: *Proteomics*.

Galaxy Proteomics is a multiple-omics' data-analysis platform with particular emphasis on mass spectrometry-based proteomics. It is hosted by GalaxyEU and maintained by the [GalaxyP](#) team from the University of Minnesota, USA.

#### GalaxyP Resources:

- About Us
- Publications
- Conference presentations
- Workshops

#### Tweets by @usegalaxy

The GalaxyP Project Retweeted

[@MetaP\\_init](#) Hello Twitter! We are the Metaproteomics Initiative, a brand new @EuPAProteomics Initiative that promotes the dissemination of #metaproteomics and its applications through collaborative networking in #microbiome studies. Follow us to stay tuned! (metaproteomics.org)

Embed View on Twitter

OPEN CHAT



# 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!*



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

Pratik Jagtap

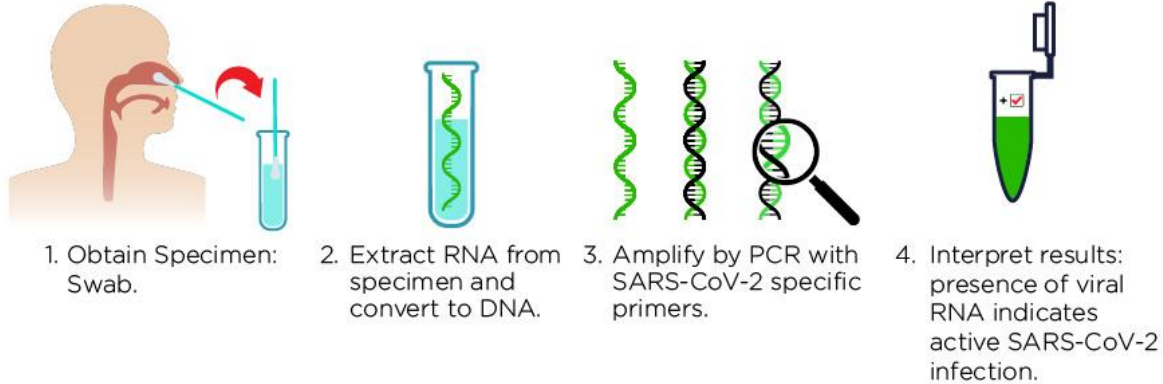
*University of Minnesota*

*Galaxy-P Team*

# COVID-19 DETECTION METHODS

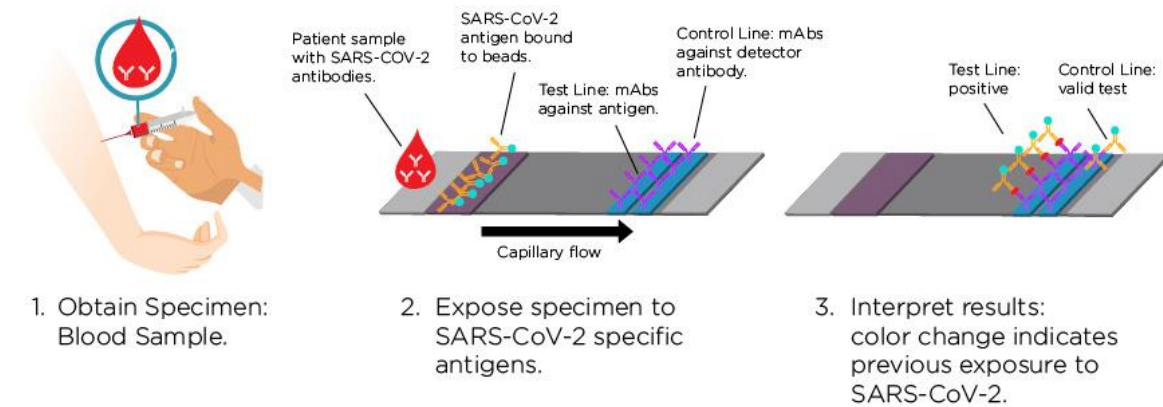
## Molecular Tests (Nucleic Acid Detection)

Diagnose active SARS-CoV-2 infections



## Antibody Tests (Serology)

Detect immune response to SARS-CoV-2 exposure



[asms.org](https://asms.org)

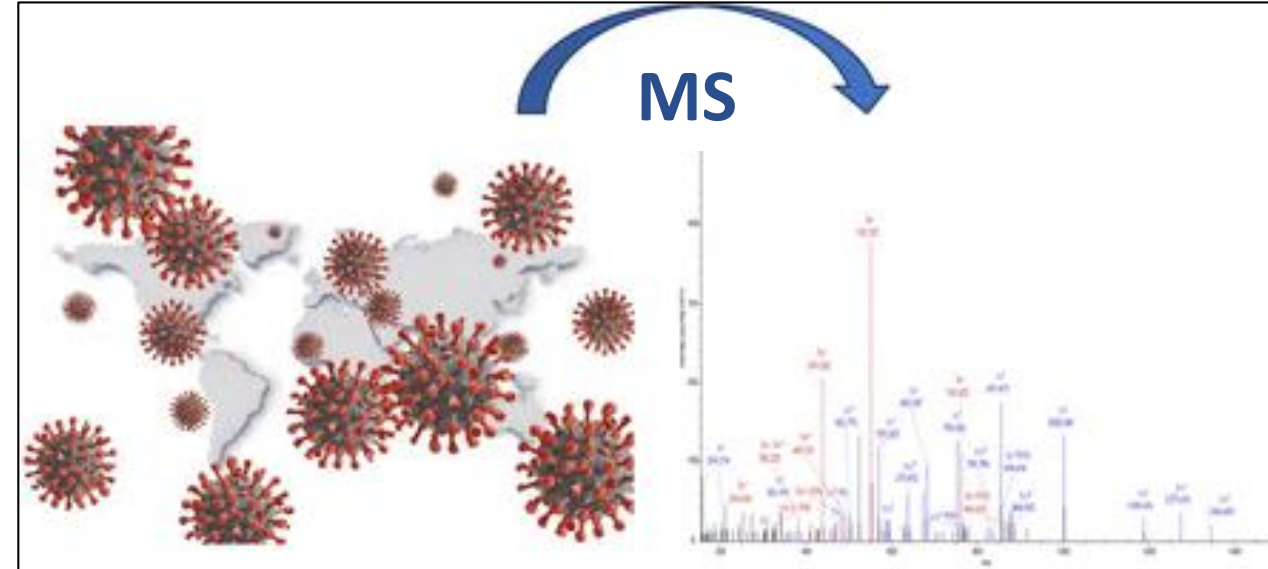














Image credit (left): Gerd Altmann, Pixabay License, <https://pixabay.com/illustrations/corona-coronavirus-virus-covid-19-4959447>




GalaxyP  
galaxyp.org

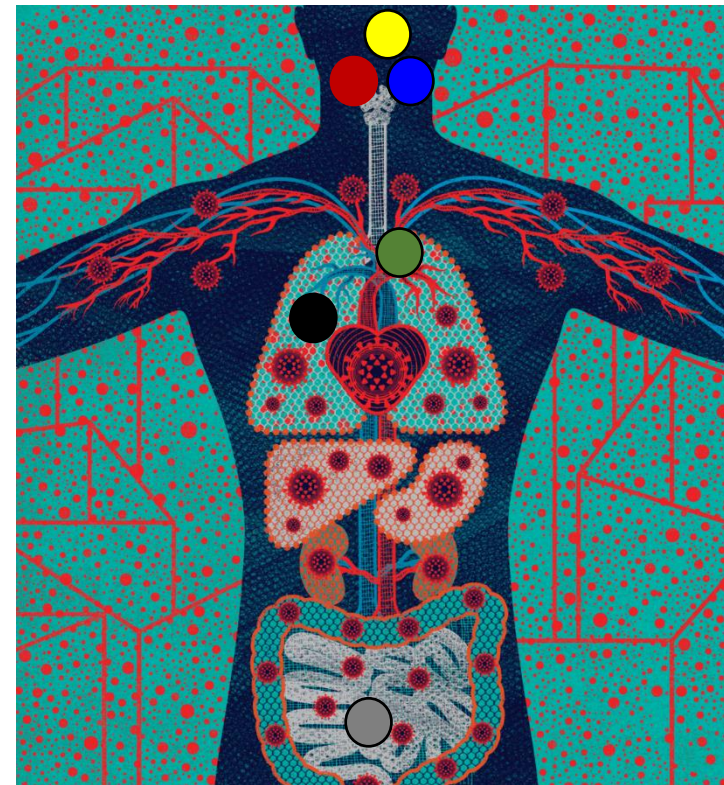


## CLINICAL SAMPLES

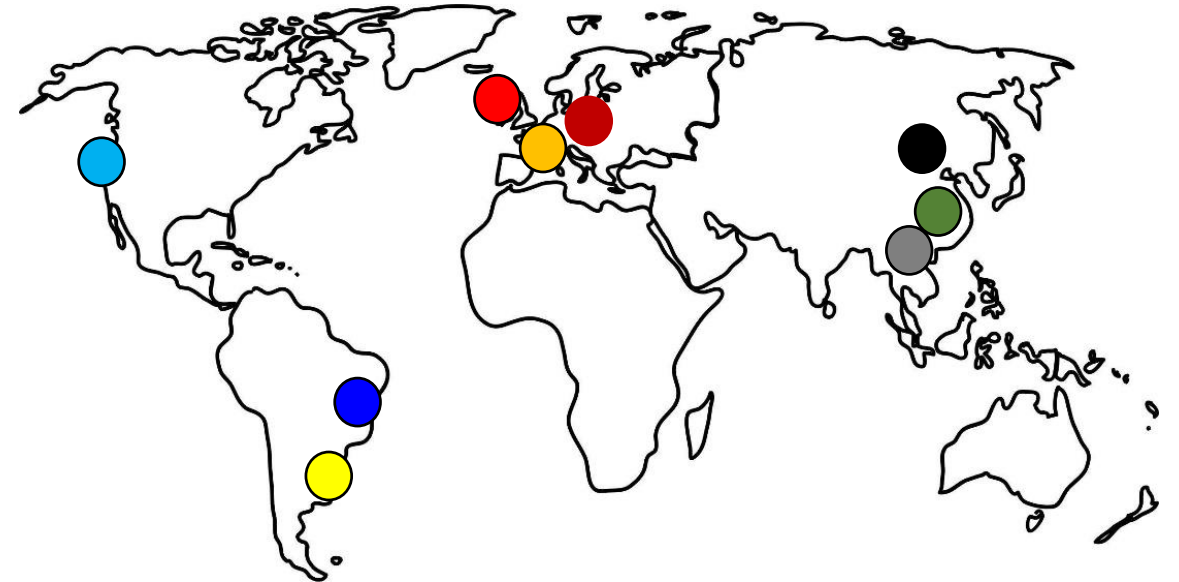
Dataset	ProteomeXchange ID   Pubmed ID	Lab
Gargling Solution 	PXD019423   PMID: 32568543	Sinz Lab (Halle, Germany) 
Nasopharyngeal swabs 	PXD020394   PMID: 32835036	Lima Lab (Montevideo, Uruguay) 
Respiratory tract samples 	PXD021328   PMID: 33273458	Carvalho Lab (São Paulo, Brazil) 
Broncho-alveolar 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 Francisco, CA) 

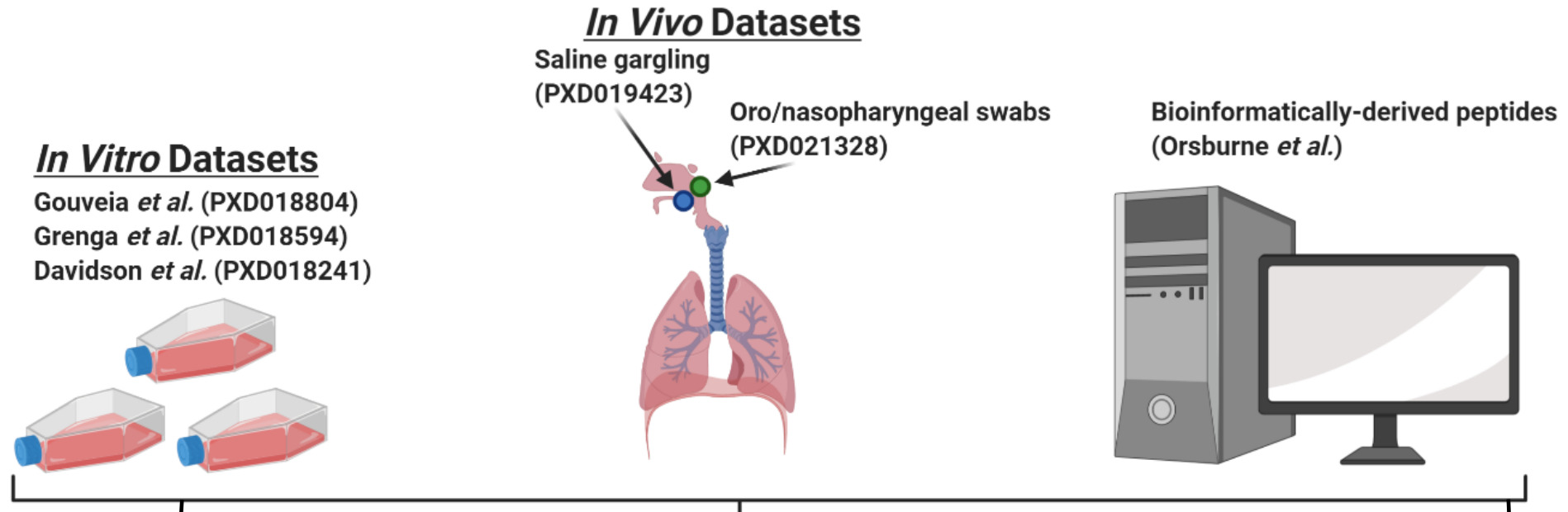


<https://www.ucsf.edu/magazine/covid-body>





# COVID-19 DETECTION MASS SPECTROMETRY METHODS



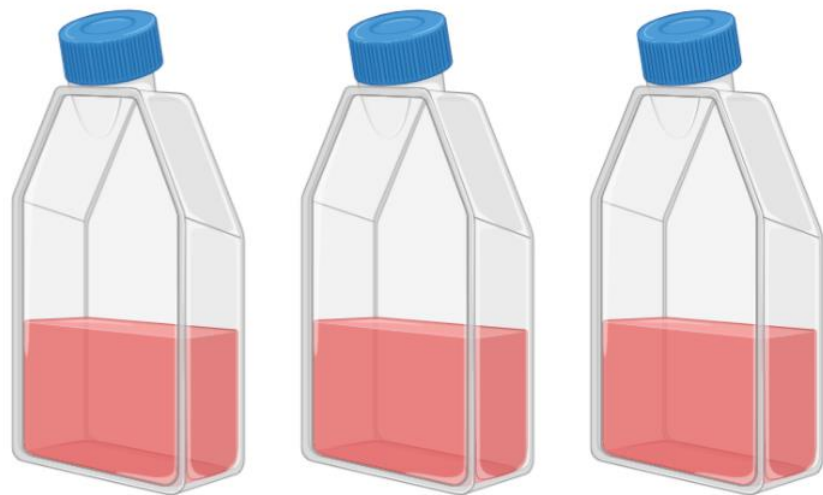
Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines. Grenga *et al*; Emerg Microbes Infect 9(1):1712-1721. doi:10.1080/22221751.2020.179173  
**7. May 2020**

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**

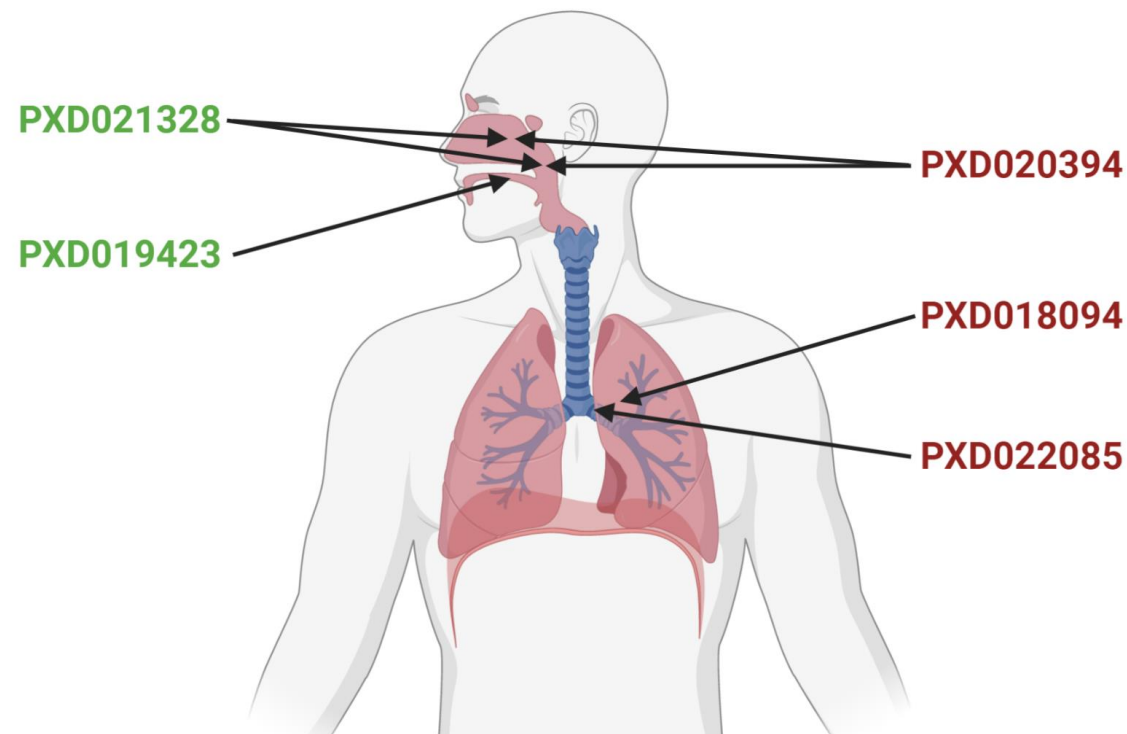
Multiple datasets were used in the creation of a peptide panel and the validation of their utility in diagnosing SARS-CoV-2

PXD018804, PXD018594, PXD018241



Peptide Panel Generation

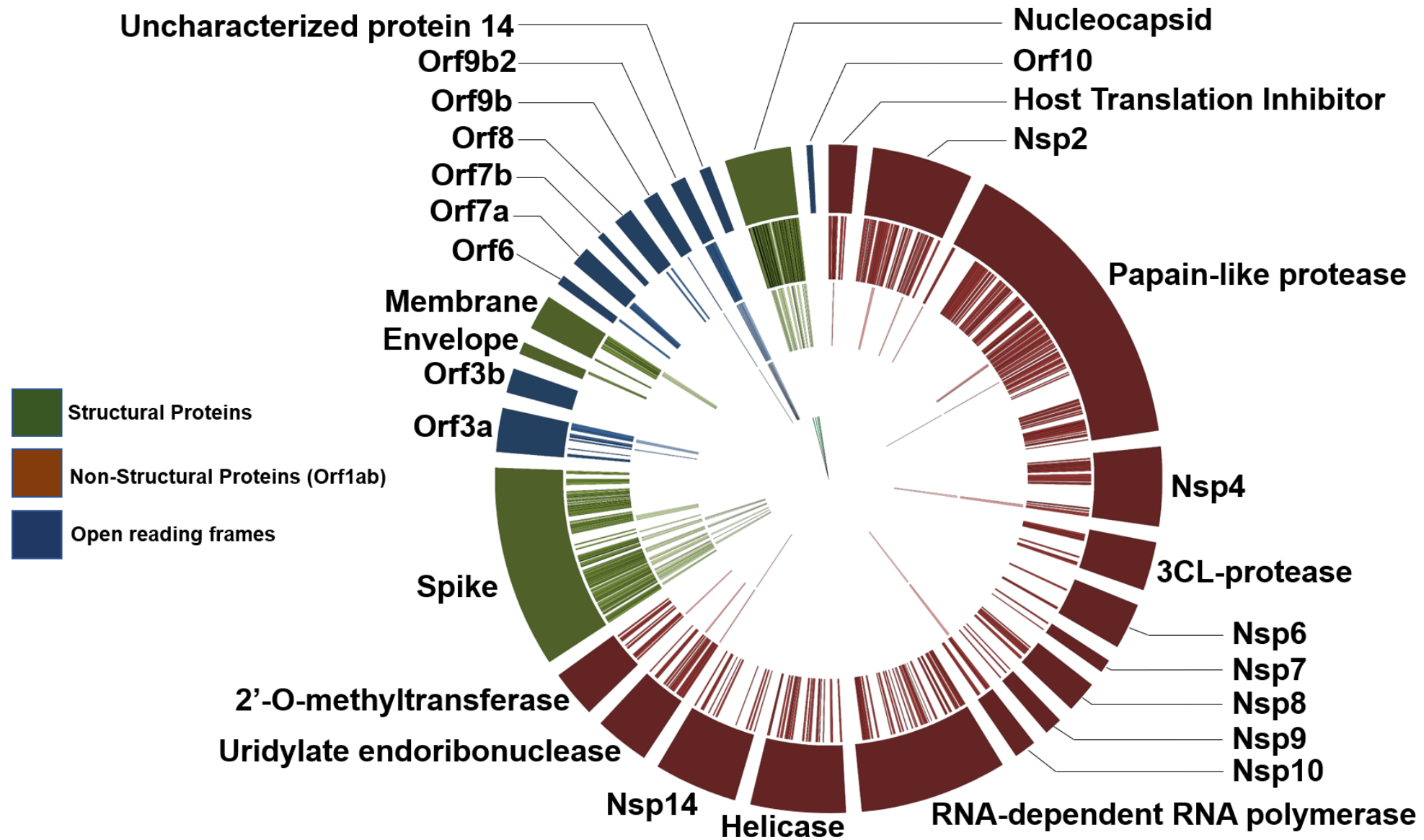
**DATABASE SEARCH WORKFLOW**



Peptide Validation

**PEPTIDE VALIDATION WORKFLOW**

# PROTEIN ASSIGNMENT OF DETECTED AND VALIDATED SARS-CoV-2 PEPTIDES

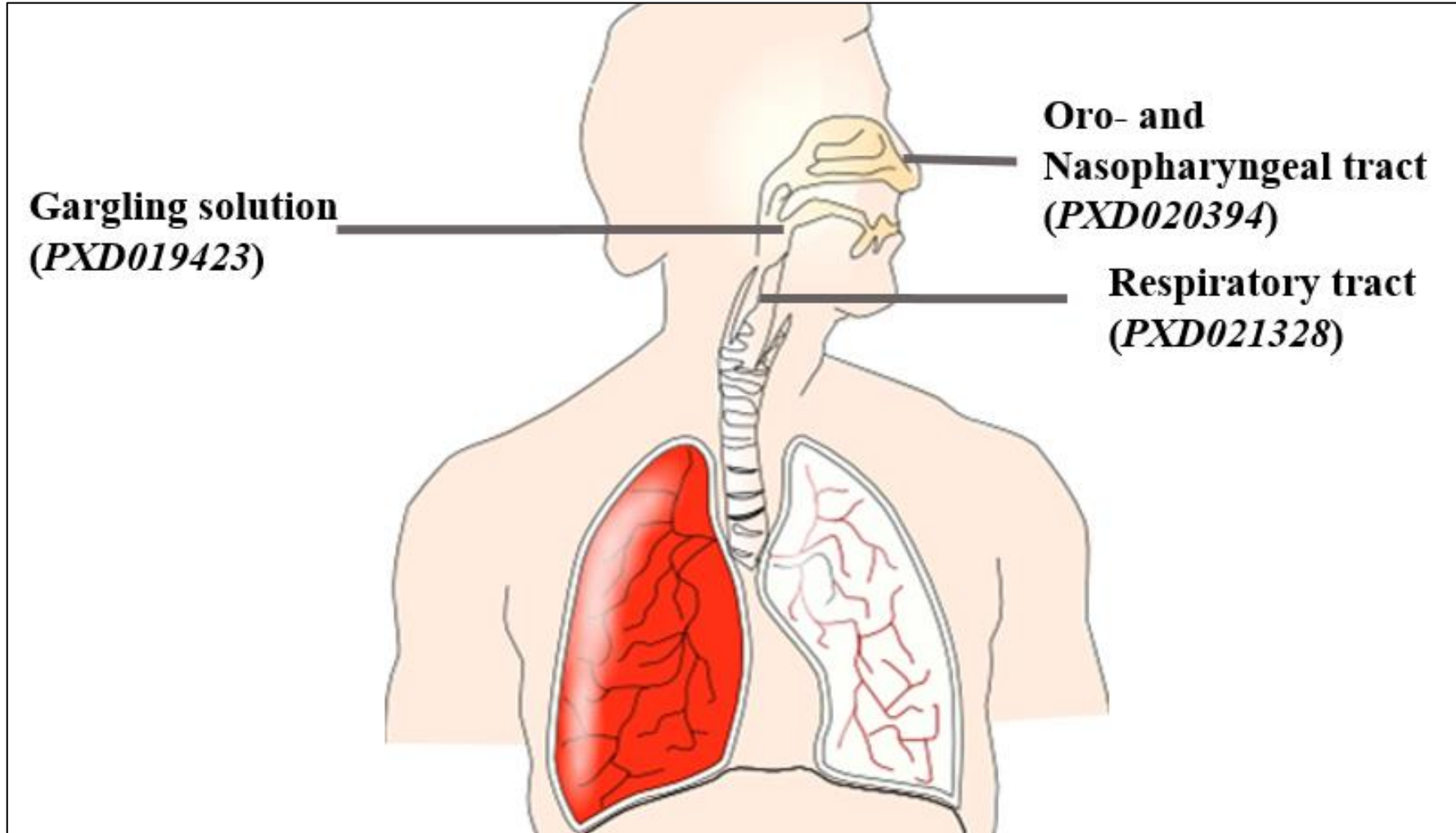


Rajczewski *et al* (2021) Accepted for publication in Clin. Proteomics.

MedRxiv Preprint: <https://www.medrxiv.org/content/10.1101/2021.02.09.21251427v2>



# 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

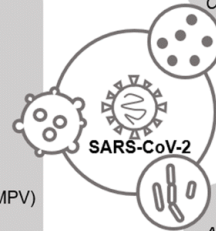
Zhu *et al* Co-infection with respiratory pathogens among COVID-2019 cases. *Virus Res.* (2020) 285:198005.

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.

**Co-infected virus**  
 Chlamydia pneumoniae  
 Coronavirus (non-COVID-19)  
 Coronavirus HKU1 (HKU1)  
 Entero/Rhinovirus (hRV)  
 H1N1  
 H3N2  
 Human metapneumovirus (hMPV)  
 Influenza A  
 Metapneumovirus  
 Mycoplasma pneumoniae  
 Parainfluenza 1/2/3/4  
 Respiratory syncytial virus (RSV)



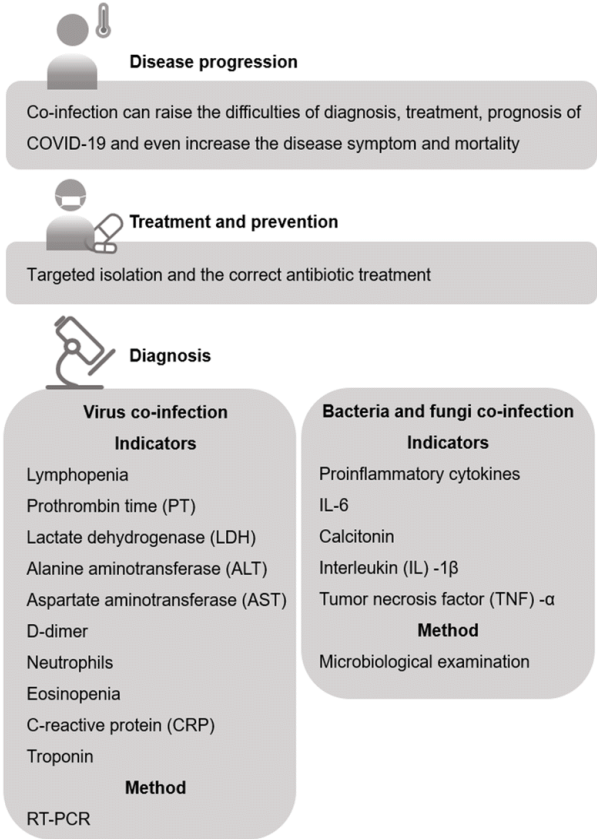
**Co-infected fungi**  
*Aspergillus spp.*  
*Candida albicans*  
*Candida glabrata*  
*Candida dubliniensis*  
*Candida parapsilosis sensu stricto*  
*Candida tropicalis*  
*Candida krusei*

**Co-infected bacteria**  
*Acinetobacter baumannii*  
*Actinomyces spp.*  
*Klebsiella pneumoniae*  
*Legionella pneumophila*  
*Rothia spp.*  
*Streptococcus spp.*  
*Veillonella spp.*



Microbial co-infection in COVID-19

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



The influence of co-infection

- 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.

# DATASETS AND ORGANISMS DETECTED

<b><i>Streptococcus pneumoniae</i></b>	Causes pneumonia (respiratory-tract infection)
<b><i>Lactobacillus rhamnosus</i></b>	Probiotic
<b><i>Pseudomonas sp. BcH</i></b>	Unclassified <i>Pseudomonas</i>
<b><i>Pseudomonas monteilii</i></b>	Meningoencephalitis
<b><i>Acinetobacter ursingii</i></b>	Bacteremia

Journal of proteome research  
pubs.acs.org/jpr Letter

1 **Metaproteomics Analysis of SARS-CoV-2-Infected Patient Samples Reveals Presence of Potential Coinfecting Microorganisms**

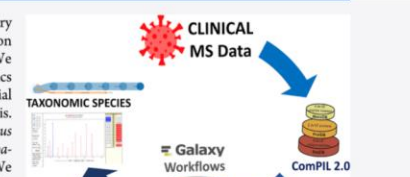
2

3 Peter S. Thuy-Boun,<sup>#</sup> Subina Mehta,<sup>#</sup> Bjoern Gruening, Thomas McGowan, An Nguyen,  
4 Andrew Rajczewski, James. E Johnson, Timothy J. Griffin, Dennis W. Wolan, and Pratik D. Jagtap\*

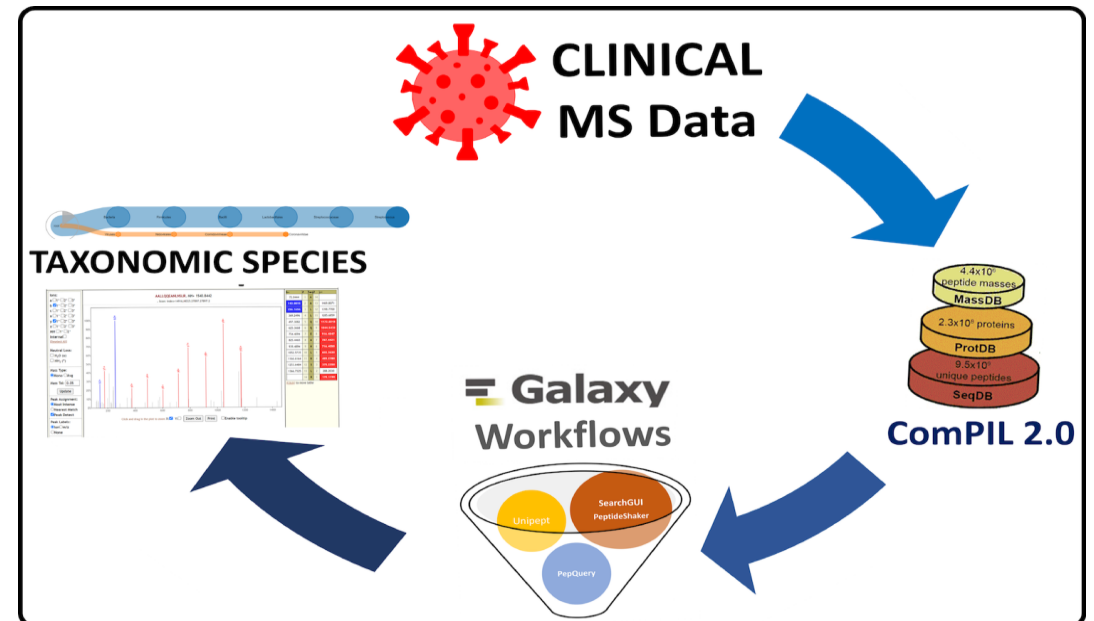
Cite This: <https://dx.doi.org/10.1021/acs.jproteome.0c00822> Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

5 **ABSTRACT:** In this Letter, we reanalyze published mass spectrometry  
6 data sets of clinical samples with a focus on determining the coinfection  
7 status of individuals infected with SARS-CoV-2 coronavirus. We  
8 demonstrate the use of ComPIL 2.0 software along with a metaproteomics  
9 workflow within the Galaxy platform to detect cohabitating potential  
10 pathogens in COVID-19 patients using mass spectrometry-based analysis.  
11 From a sample collected from gargling solutions, we detected *Streptococcus*  
12 *pneumoniae* (opportunistic and multidrug-resistant pathogen) and *Lactoba-*  
13 *cillus rhamnosus* (a probiotic component) along with SARS-Cov-2. We



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](#)

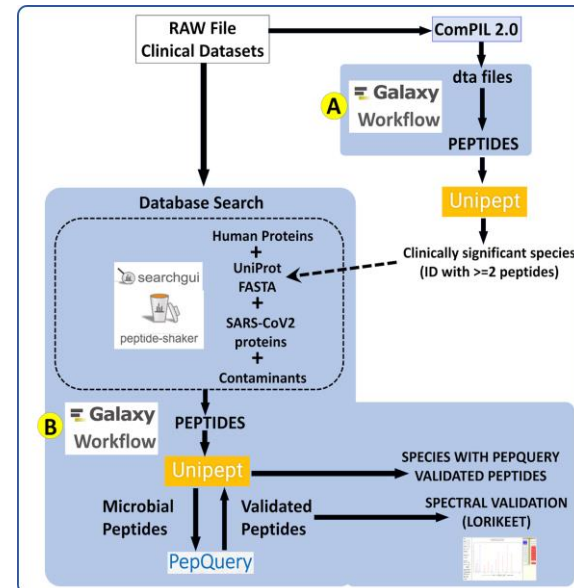
• [Metaproteomics of mPXD020394](#)

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>



# 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](https://usegalaxy.eu).

Once Registered, click on <https://usegalaxy.eu/join-training/glbio2021> to join the GLBIO 2021 Galaxy session.

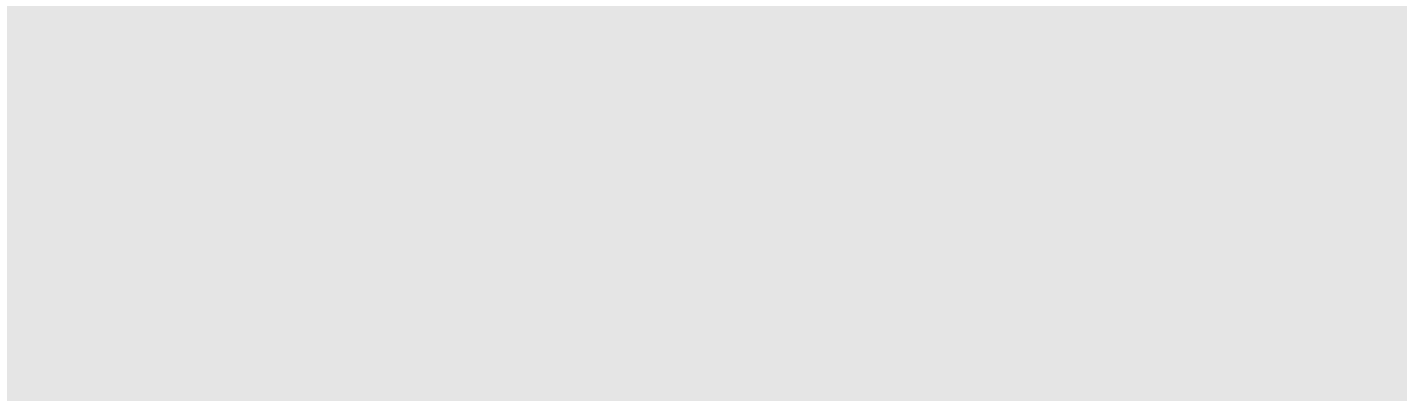
Go to Shared Data  
→ Histories  
→ GLBIO inputs  
Import +

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

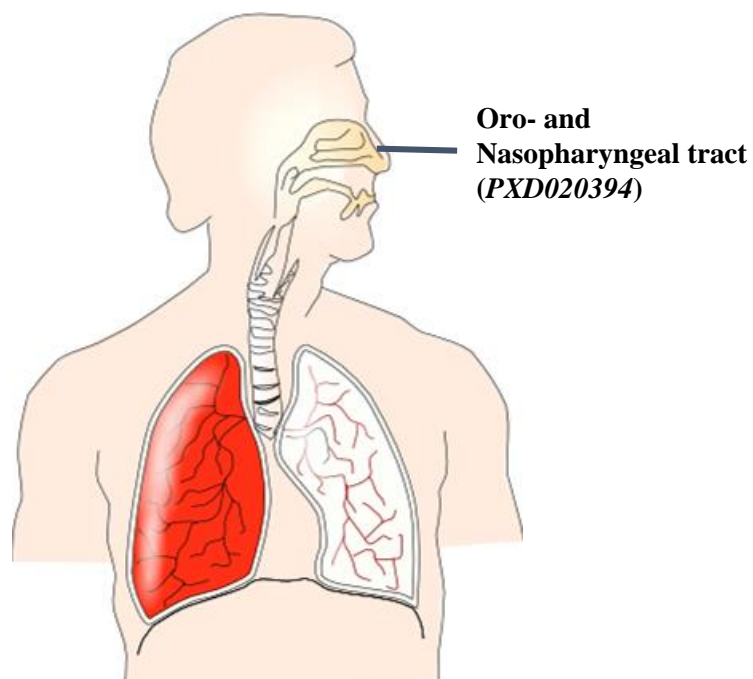
Run the workflow on active history



# Determining the presence of COVID-19 peptides in Galaxy



# Dataset Information



Oro-Nasopharyngeal

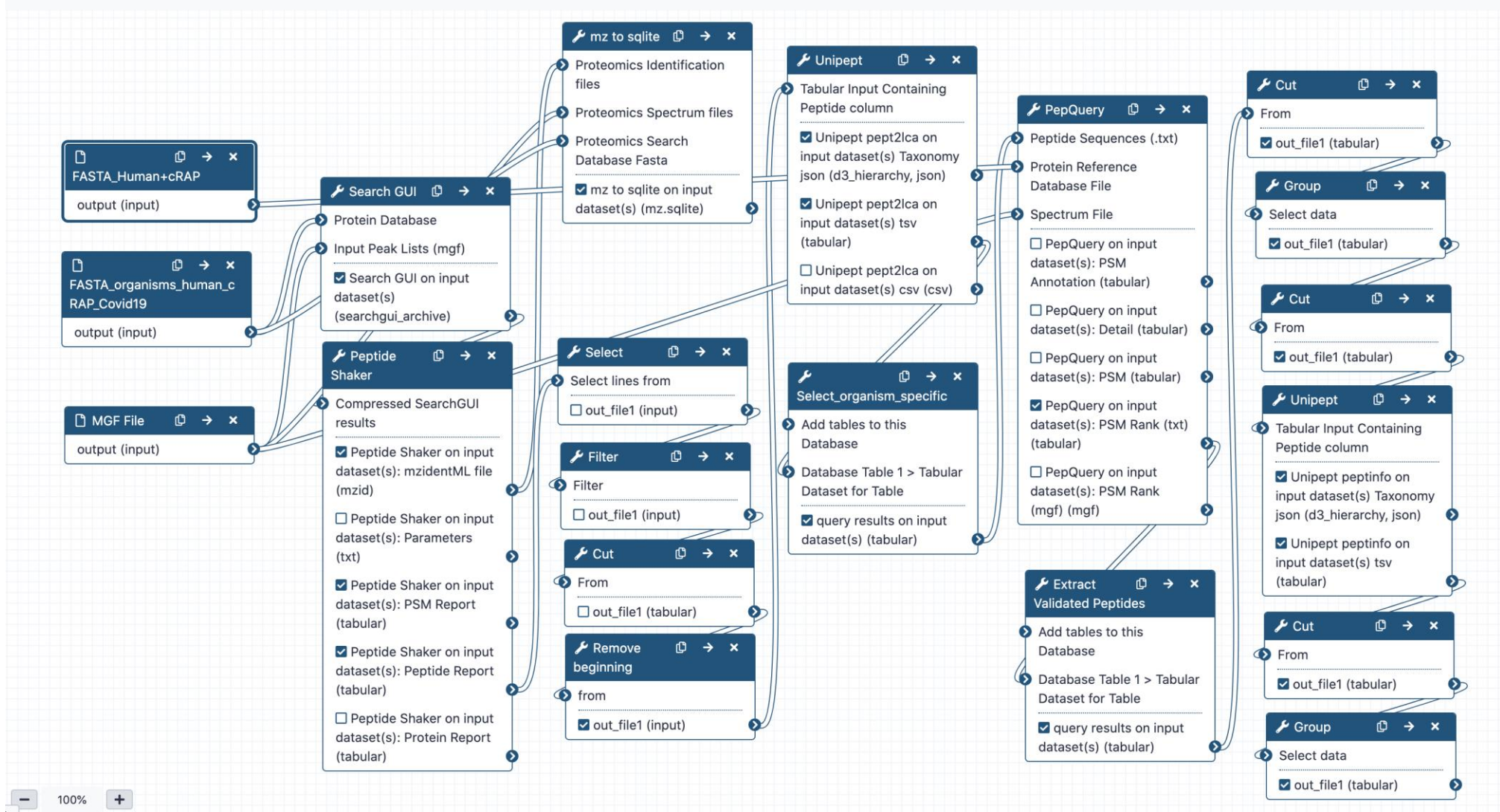
Patient data - Positive and Negative

Coronavirus present?

Any other species – co-infection?

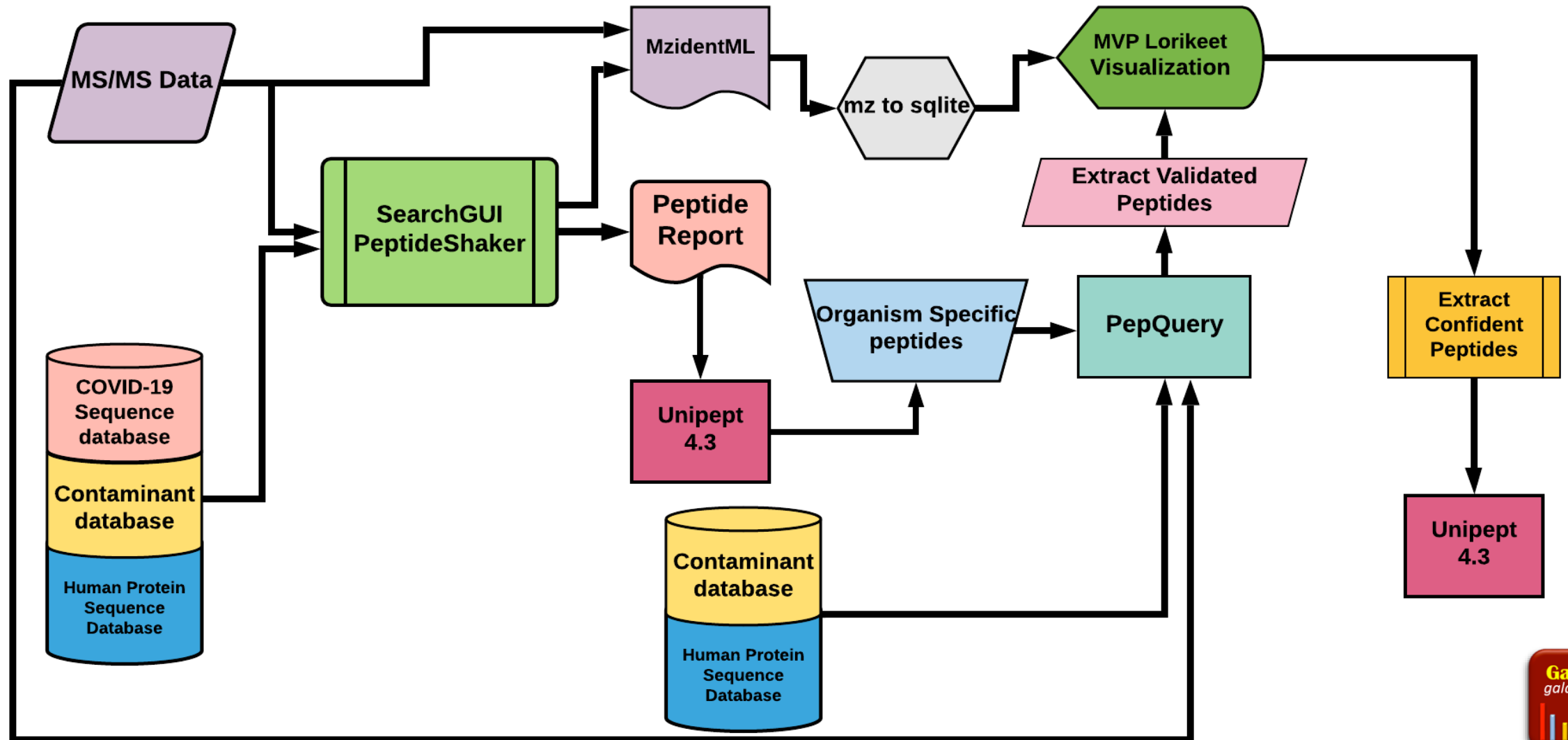
# GALAXY WORKFLOW

GLBIO workflow





# 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

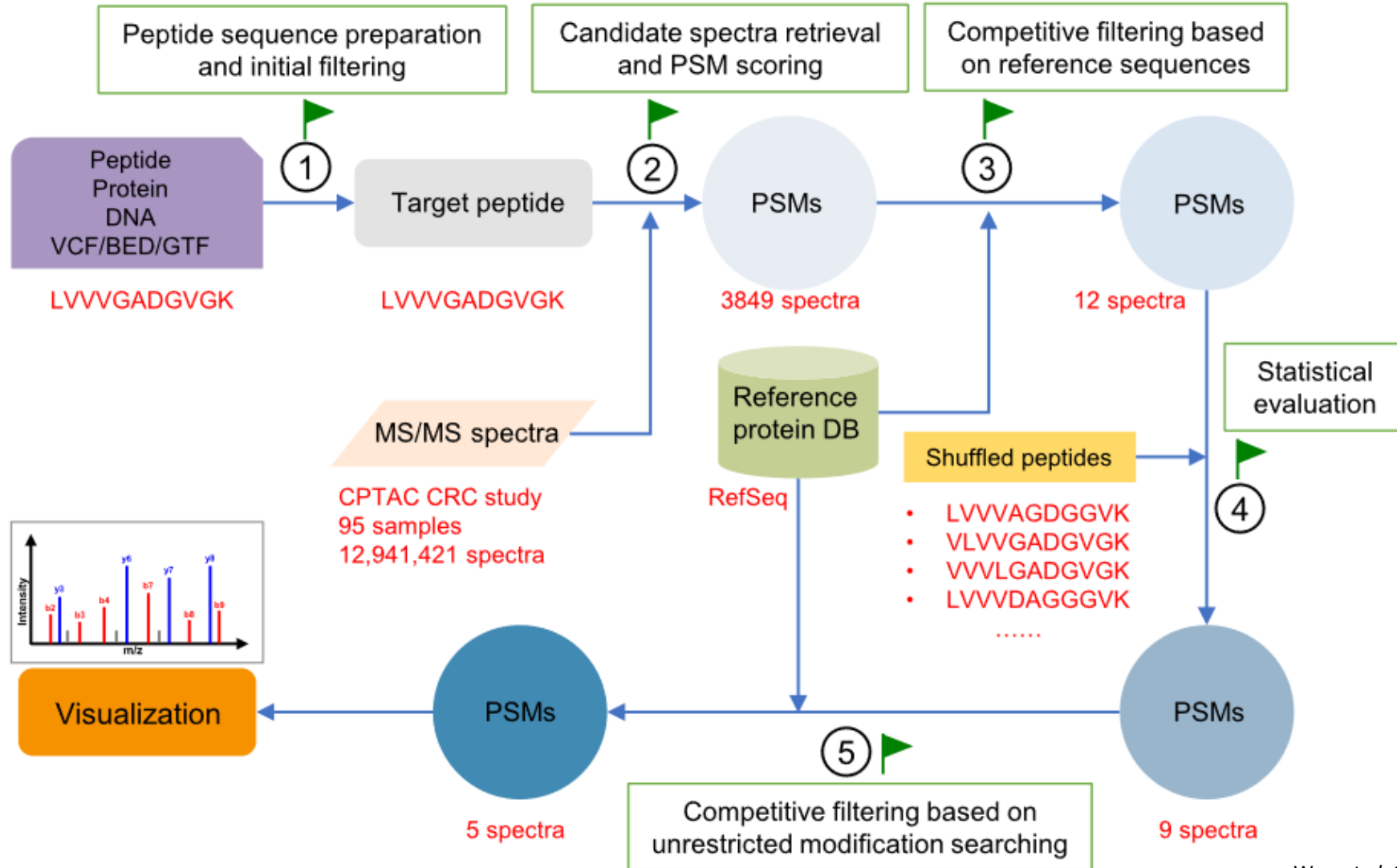
select organisms  
with at least 2  
peptides

# Peptide Validation

PepQuery

Spectral Quality

# PepQuery Workflow



# PepQuery

Features  
PSM rank

Eliminate peptides  
in reference  
database

P-value

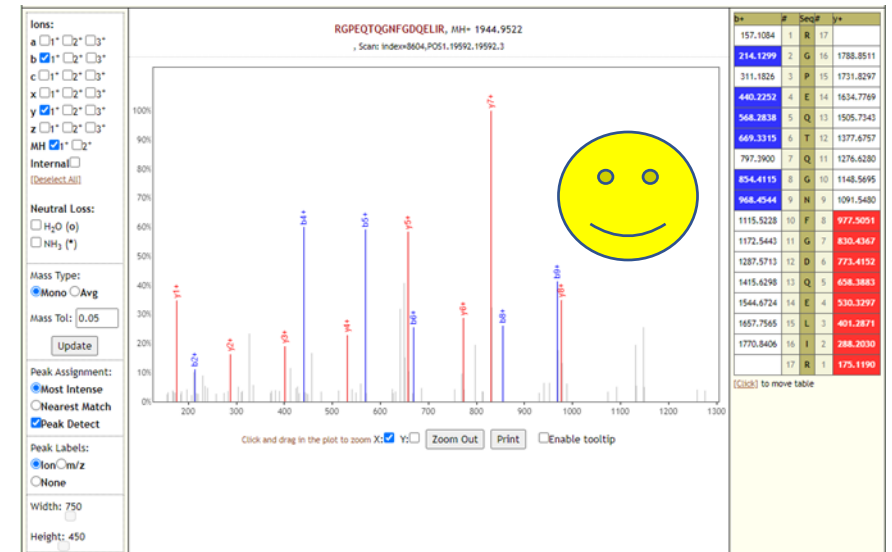
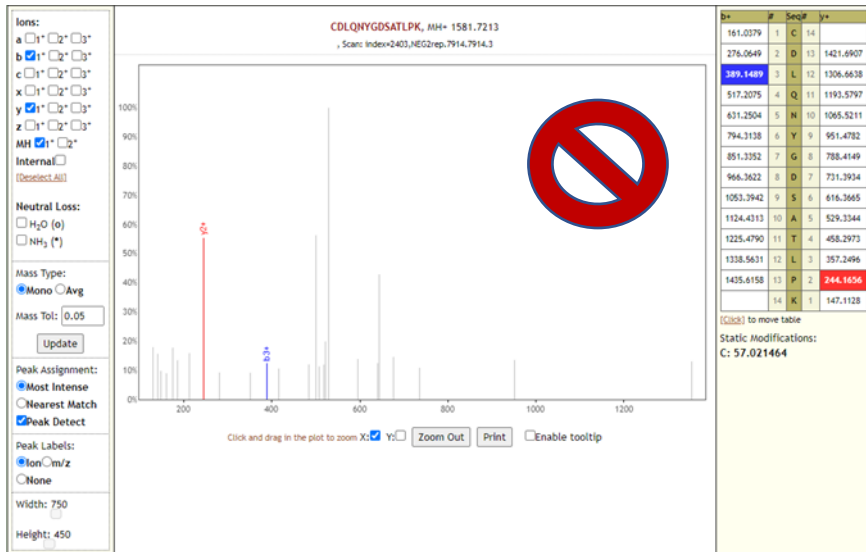
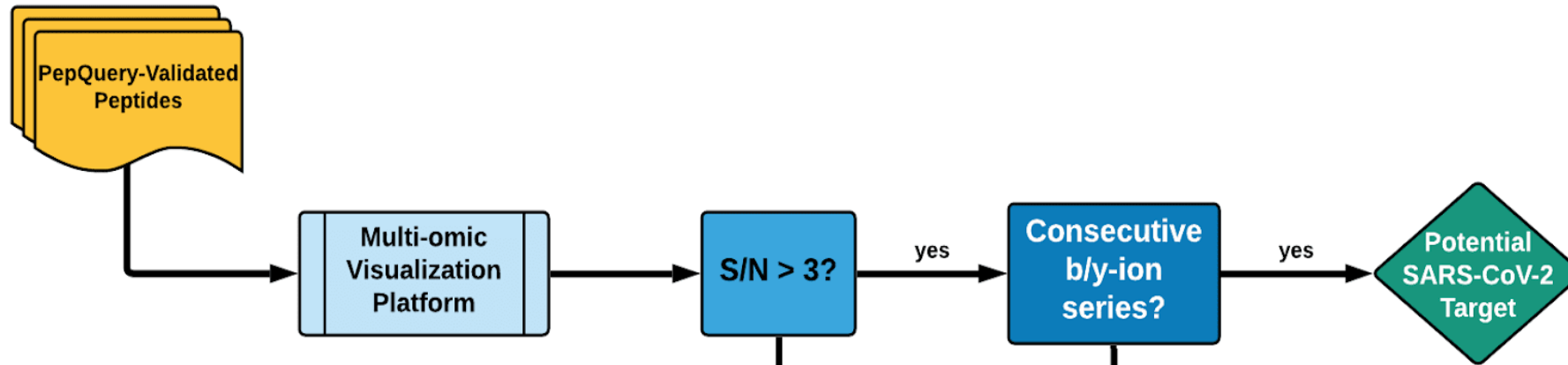
confidence

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](http://galaxyp.org)
- [proteomics.usegalaxy.eu](http://proteomics.usegalaxy.eu)
- [Galaxy Workflows for COVID-19 Research](#)
- [Galaxy Training Network](#)
- [Co-infection Metaproteomics](#)
- [Peptides for clinical diagnosis of COVID-19](#)