Galaxy Workflows for analysis of COVID-19 Mass Spectrometry datasets

24 February 2021
10.00 CT/ 11.00 ET/ 17.00 CET

Pratik Jagtap
Subina Mehta
University of Minnesota
Galaxy-P Team
Housekeeping

Please remain muted unless you’re invited to speak by the Chair.

Please use the “hand-raising function” to indicate you would like to contribute directly.

This meeting will be run in line with the ELIXIR Code of Conduct. If you have any concerns please refer to the Code of Conduct, found on the ELIXIR website.

Please use “Q&A” to raise questions during the presentation.

Please use “Chat” for further comments or discussions.

This session will be recorded.
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.
**COVID-19 DETECTION MASS SPECTROMETRY METHODS**

**In Vivo Datasets**
- Gouveia *et al.* (PXD018804)
- Grenga *et al.* (PXD018594)
- Davidson *et al.* (PXD018241)

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

**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](https://doi.org/10.1101/2020.03.08.980383) April 2020

**Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines.**

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Unpublished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

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ProteomeXchange ID</th>
<th>Pubmed ID</th>
<th>Lab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time series</td>
<td>PXD018594</td>
<td>PMID: 32619390</td>
<td>Armengaud Lab (Bagnols-sur-Cèze, France)</td>
</tr>
<tr>
<td>8 hours time point</td>
<td>PXD018804</td>
<td>PMID: 32462744</td>
<td>Armengaud Lab (Bagnols-sur-Cèze, France)</td>
</tr>
<tr>
<td>Proteo-transcriptomic analysis</td>
<td>PXD018241</td>
<td>PMID: 32723359</td>
<td>Matthews Lab (Bristol, UK)</td>
</tr>
<tr>
<td>Host-viral protein interaction</td>
<td>PXD018117</td>
<td>PMID: 32353859</td>
<td>Krogan Lab (San Francisco, CA)</td>
</tr>
</tbody>
</table>
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

https://covid19.galaxyproject.org/proteomics

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

Peter Thuy-Boun et al
http://dx.doi.org/10.1021/acs.jproteome.0c00822
Determining the optimal peptides for COVID-19 diagnosis in Galaxy
Multiple datasets were used in the creation of a peptide panel and the validation of their utility in diagnosing SARS-CoV-2
Database Search Workflow
Peptide Validation Workflow

1. SARS-Related Protein Sequence DB and Contaminant database
2. PepQuery
3. 639 COVID-19 Peptides
4. Extract Confident Peptides
5. MVP Lorikeet Visualization
6. Final Peptide List
7. Peptide Shaker Output
8. mz to sqlite
9. Blast-P NR
10. Unipept 4.3
Peptides across SARS-CoV-2 were detected and validated in Galaxy

<table>
<thead>
<tr>
<th>Color</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Green</td>
<td>Papain like proteinase</td>
</tr>
<tr>
<td>Yellow</td>
<td>Uriddylate-specific endoribonuclease</td>
</tr>
<tr>
<td>Orange</td>
<td>Non-structural protein</td>
</tr>
<tr>
<td>Purple</td>
<td>2'-O-methyltransferase</td>
</tr>
<tr>
<td>Turquoise</td>
<td>Host translation inhibitor</td>
</tr>
</tbody>
</table>

![Diagram of viral proteins and peptides]

Membrane (M) Protein
Spike (S) Protein
Nucleocapsid (N) and other viral RNA Proteins

- Peptide sequences:
  - NUCLEOCAPSID-PHOSPHOPROTEIN
  - ORF 3a
  - SPIKE PROTEIN

- Accession numbers:
  - PXD021328
  - PXD019423
  - PXD020493
PSMs of SARS-CoV-2 peptides in the upper respiratory clinical datasets are of higher confidence than deep lung datasets.
Protein assignment of detected and validated SARS-CoV-2 peptides
Four peptides were selected as optimal targets for SARS-CoV-2 detection.
BLAST-P shows specificity of these peptides to SARS-CoV-2.
Conclusions

• Based on Clinical and Cell culture MS datasets, we identified peptides throughout the SARS-CoV-2 proteome.

• High-confidence PepQuery scoring and manual spectra interrogation reveal four confident peptides.
  - MAGNGGDAALALLLLDR
  - RGPEQTQGNFGDQELIR
  - DGIIWVATEGALNTPK
  - IGMEVTPSGTWLTYTGAIK

• Deep-lung samples may be unsuitable for diagnosis of COVID-19 using targeted clinical proteomics experiments.
Metaproteomics analysis of SARS-CoV-2-infected patient samples reveals presence of potential co-infecting microorganisms.
Co-infection in COVID-19 Patients

- 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.


Metaproteomics Workflow

Oropharyngal and Nasopharyngeal tract (PXD019423)
Respiratory tract (PXD021328)

**ABSTRACT.** In this Letter, we analyze published mass spectrometry data sets of clinical samples with a focus on determining the co-infection status of individuals infected with SARS-CoV-2 coronavirus. We demonstrate the use of ComPIL 2.0 software along with a metaproteomics workflow within the Galaxy platform to detect co-infecting potential pathogens in COVID-19 patients using mass spectrometry-based analysis.

From a sample collected from gargling solution, we detected Streptococcus pneumoniae (opportunistic and multiresistant pathogen) and Lactobacillus rhamnosus (a probiotic component) along with SARS-CoV-2. We could also detect Pseudomonas spp. from COVID-19 positive samples and Acinetobacter spp. and Pseudomonas aeruginosa from COVID-19 negative samples from oro- and nasopharyngeal samples. We believe that the early detection and characterization of co-infections by using metaproteomics from COVID-19 patients will potentially impact the diagnosis and treatment of patients infected with SARS-CoV-2.

https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00822
### Datasets and Organisms Detected

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Organisms detected in COVID-19 patient samples</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gargling solution (PXD019423)</td>
<td><em>Streptococcus pneumoniae</em>, <em>Lactobacillus rhamnosus</em> and SARS-CoV-2</td>
<td><a href="https://covid19.galaxyproject.org/proteomics/mPXD019423/">https://covid19.galaxyproject.org/proteomics/mPXD019423/</a></td>
</tr>
</tbody>
</table>

#### Organisms and Their Properties

<table>
<thead>
<tr>
<th>Organism</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Streptococcus pneumoniae</em></td>
<td>Causes pneumonia (respiratory-tract infection)</td>
</tr>
<tr>
<td><em>Lactobacillus rhamnosus</em></td>
<td>Probiotic</td>
</tr>
<tr>
<td><em>Pseudomonas monteilii</em></td>
<td>Unclassified <em>Pseudomonas</em></td>
</tr>
<tr>
<td><em>Acinetobacter ursingii</em></td>
<td>Bacterimia</td>
</tr>
</tbody>
</table>
Spectral Validation (Lorikeet)

Lactobacillus rhamnosus

SARS CoV-2

Acinetobacter ursingii

Streptococcus pneumoniae
Conclusions

• Galaxy workflows are available for analysis of COVID-19 MS datasets. (https://covid19.galaxyproject.org/proteomics/)

• 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.

• Future plans
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

Resources available at https://covid19.galaxyproject.org/proteomics
ACKNOWLEDGMENTS

University of Minnesota
Timothy Griffin
Subina Mehta
Andrew Rajczewski
Dinh Duy An Nguyen
Emma Leith
Ray Sajulga
Praveen Kumar
Caleb Easterly
Marie Crane

Minnesota Supercomputing Institute
James Johnson
Michael Milligan

Brian Searle
Institute of Systems Biology

Brook Nunn
U of Washington

Peter Thuy-Boun
Dennis Wolan
Scripps Institute

Lloyd Smith (Co-I)
Michael Shortreed
UW-Madison

Jeremy Fisher
Yuzhen Ye
Sujun Li
Indiana University

Matt Chambers
Nashville, TN

GalaxyP

Saskia Hiltemann

Lennart Martens
Bart Mesuere
VIB, UGhent, Belgium

Alessandro Tanca
Porto Conte Ricerche, Italy

Bjoern Gruening
University of Freiburg, Freiburg, Germany

Thilo Muth
Robert Koch Institut

Anamika Krishanpal
Persistent Systems Limited

Magnus Øverlie Arntzen
NMBU, Oslo, Norway

Carolin Kolmeder
University of Helsinki, Finland

Biologists / collaborators
Joel Rudney
Maneesh Bhargava
Amy Skubitz
Chris Wendt
Kristin Boylan
Brian Sandri
Alexa Pragman

Funding

Twitter: twitter.com/usegalaxyp

Website: galaxyp.org
To get involved: https://galaxyproject.org/community/
Training materials: https://training.galaxyproject.org/
Further information: https://usegalaxy.org/
3rd Galaxy webinar series: Advanced Features

3 March: Advanced Workflow features
10 March: Processing thousands of datasets simultaneously
17 March: Bridging two worlds (Jupyter Notebooks and RStudio)
24 March: Features no-one knows about