A rigorous evaluation of optimal peptide targets for MS-based clinical diagnostics of Coronavirus Disease 2019 (COVID-19)

Andrew T. Rajczewski[1]; Subina Mehta[1]; Dinh Duy An Nguyen[1]; Björn Grüning[2]; James E. Johnson[1]; Thomas McGowan[1]; Timothy J. Griffin[1]; Pratik D. Jagtap[1]

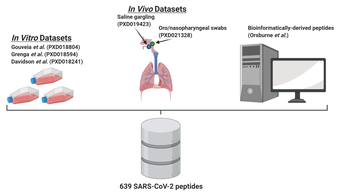
[1] Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN

[2] Department of Computer Science, University of Freiburg, Freiburg im Breisgau, Germany

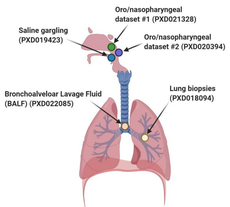
INTRODUCTION

- The COVID-19 pandemic has wrought immense suffering worldwide, with over 110 million cases and nearly 2.5 million deaths as of February 2021
- Rapid diagnosis of SARS-CoV-2 infection is the key to identification and early treatment of patients and minimizing the spread of the disease
- Targeted mass spectrometry (MS) represents a potential complementary mode of detecting SARS-CoV-2 that doesn't rely on amplification of the viral RNA
- To maximize the ability of targeted MS assays to detect and quantify SARS-CoV-2, target peptides should be selected for detection that have the greatest likelihood of being reliably observed in patient samples
- A selection of COVID19 datasets were analyzed to build a panel of potential SARS-CoV-2 peptide targets
- SARS-CoV-2 panel was assessed against COVID19 datasets to determine the quality of peptide spectral matches
- Four peptides were selected as the best candidate target for MS-based detection of SARS-CoV-2 in patient samples

EXPERIMENTAL METHODS

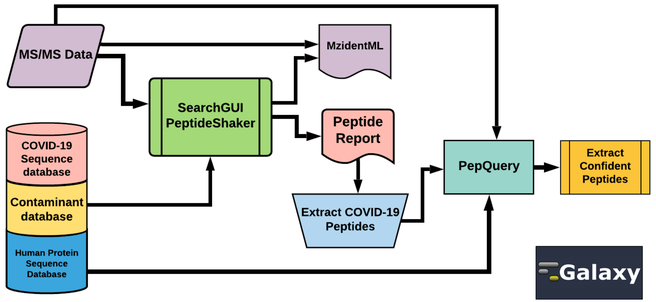


- Cell culture experiments and analyzed patient samples were searched against SARS-CoV-2 proteome in a bioinformatics workflow in Galaxy (right)
- Resulting data were combined with theoretical peptides to generate our panel of potential targets



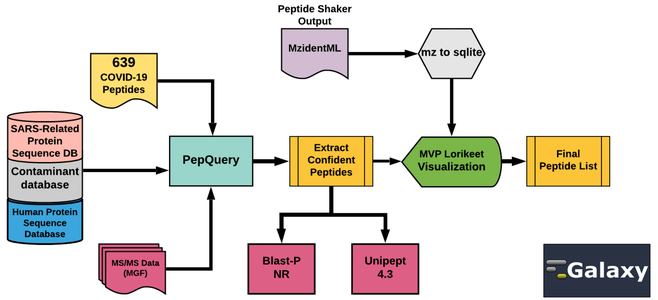
- Our peptide panel was searched against datasets from pooled oropharyngeal and nasopharyngreal swabs, saline gargling, lung biopsy tissue, and bronchoalveolar lavage fluid (BALF)
- Our bioinformatics workflow was able to assess the quality of peptide spectra in Galaxy (right)

GENERATION OF SARS-COV-2 PEPTIDE PANEL



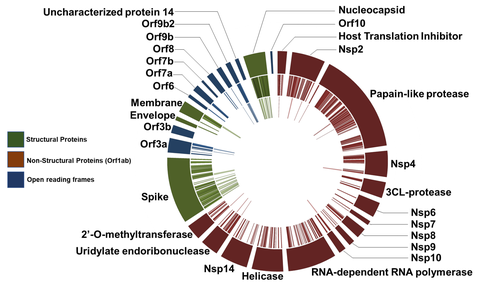
- MS/MS data from cell culture experiments and analyzed patient samples were searched against SARS-CoV-2 protein FASTA files from Uniprot
- SearchGUI/Peptide Shaker results were searched against human proteome in PepQuery and filtered to leave only viral peptides

VALIDATION OF PEPTIDES IN PATIENT DATASETS

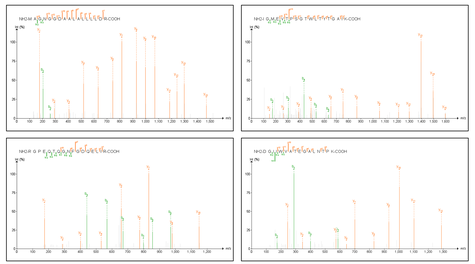


- PepQuery search engine ranks the matches of our peptide panel to raw COVID19 data against a reference proteome containing proteins from human and other coronaviruses
- Resulting data were filtered to give the most confident matches to SARS-CoV-2 and subsequently validated using bioinformatic tools such as BLAST-P and Unipept in addition to manual evaluation of spectral quality using the Multi-omics Viewing Platform (MVP)

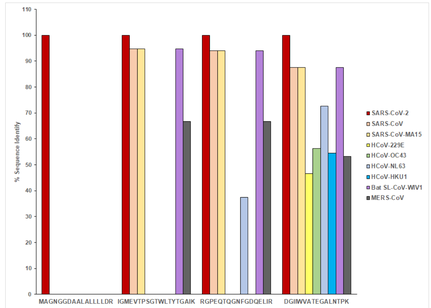
OUT OF A LARGE PANEL OF SARS-COV-2 PEPTIDES, FOUR PEPTIDES REPRESENT THE STRONGEST POTENTIAL TARGETS



- 639 peptides were found across all the cell culture and patient datasets, corresponding to most of the viral proteome (inner diameter circle)
- PepQuery validation of the peptide panel found 75 peptides that passed validation, with these peptides corresponding principally to structural proteins (Nucleocapsid, Spike, and Membrane proteins) as well as portions of some non-structural proteins and open reading frames (innermost circle)



- MS/MS spectra of peptides that passed validation through PepQuery were manually inspected to determine their completeness
  - At least 3 sequential b- or y-ions
  - Product ion S/N ratio of at least 3
- Peptides with the highest confidence match, best quality spectra, and were found in the most patient datasets were selected as being suitable targets for SARS-CoV-2 detection
  - Four peptides from the Nucleocapsid protein
  - MAGNGGDAALALLLLDR, IGMEVTPSGTWLTYTGAIK, RGPEQTQGNFGDQELIR, DGIIWVATEGALNTPK



- BLAST-P analysis of four target peptides shows sequence specificity to SARS-CoV-2 over other coronaviruses
- MAGNGGDAALALLLLDR was found to be unique to SARS-CoV-2, with no alignment to other coronaviruses
- None of the selected peptides showed any alignment to the human proteome

SUMMARY

- Peptides from SARS-CoV-2 were evaluated to determine their suitability for diagnosis of COVID-19 through targeted mass spectrometry
- Analysis of cell culture and patient datasets established a panel of 639 SARS-CoV-2 peptides which covered most of the viral proteome
- Seventy-five peptides from the panel were validated in patient data
- Four peptides from the nucleocapsid protein were found to have high confidence in their PepQuery validation, high quality spectra, and be present in most patient samples
- The selected target peptides are specific to SARS-CoV-2
- This work is available as a pre-print on Medrxiv https://www.medrxiv.org/content/10.1101/2021.02.09.21251427v1

REFERENCES

1. Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W., A Novel Coronavirus from Patients with Pneumonia in China, 2019. New England Journal of Medicine 2020, 382 (8), 727-733.
2. Tichopad, A.; Didier, A.; Pfaffl, M. W., Inhibition of real-time RT–PCR quantification due to tissue-specific contaminants. Molecular and Cellular Probes 2004, 18 (1), 45-50
3. Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M., Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proceedings of the National Academy of Sciences 2007, 104 (14), 5860-5865.
4. Wu, F.-X.; Gagné, P.; Droit, A.; Poirier, G. G., Quality assessment of peptide tandem mass spectra. BMC Bioinformatics 2008, 9 (S6), S13.
5. Wen, B.; Wang, X.; Zhang, B., PepQuery enables fast, accurate, and convenient proteomic validation of novel genomic alterations. Genome research 2019, 29 (3), 485-493.
6. McGowan, T.; Johnson, J. E.; Kumar, P.; Sajulga, R.; Mehta, S.; Jagtap, P. D.; Griffin, T. J., Multi-omics Visualization Platform: An extensible Galaxy plug-in for multi-omics data visualization and exploration. GigaScience 2020, 9 (4).

ACKNOWLEDGEMENTS

We would like to acknowledge Emma Leith for her assistance in the generation of this data. We'd also like to acknowledge funding for this work from the grant National Cancer Institute - Informatics Technology for Cancer Research (NCI-ITCR) grant 1U24CA199347 and National Science Foundation grant 1458524. The European Galaxy server that was used for data analysis is in part funded by Collaborative Research Centre 992 Medical Epigenetics (DFG grant SFB 992/1 2012) and German Federal Ministry of Education and Research (BMBF grants 031 A538A/A538C RBC, 031L0101B/031L0101C de.NBI-epi, 031L0106 de.STAIR (de.NBI)). Andrew T. Rajczewski was supported by Biotechnology Training Grant: NIH T32GM008347