

ADVANCING MASS SPECTROMETRY-BASED COVID-19 DIAGNOSTICS BY IDENTIFYING OPTIMAL BIOMARKER PEPTIDES FROM EMERGING STRAINS FROM NON-INVASIVELY COLLECTED CLINICAL SAMPLES



Subina Mehta¹, Andrew T. Rajczewski¹, Björn Grüning², James E. Johnson¹, Reid Wagner¹, Olivier Pible³, Jean Armengaud³, Valdemir M. Carvalho⁴, Timothy J. Griffin¹, Pratik D. Jagtap¹

1. University of Minnesota, Minneapolis, MN; 2. Department of Computer Science, University of Freiburg, Georges-Köhler-Allee, Freiburg, 3. CEA Life Sciences Division, DSV, IBEB, SBTN, Bagnols-sur-Cèze, France. 4. Division of Research and Development, Fleury Group, São Paulo, SP, Brazil

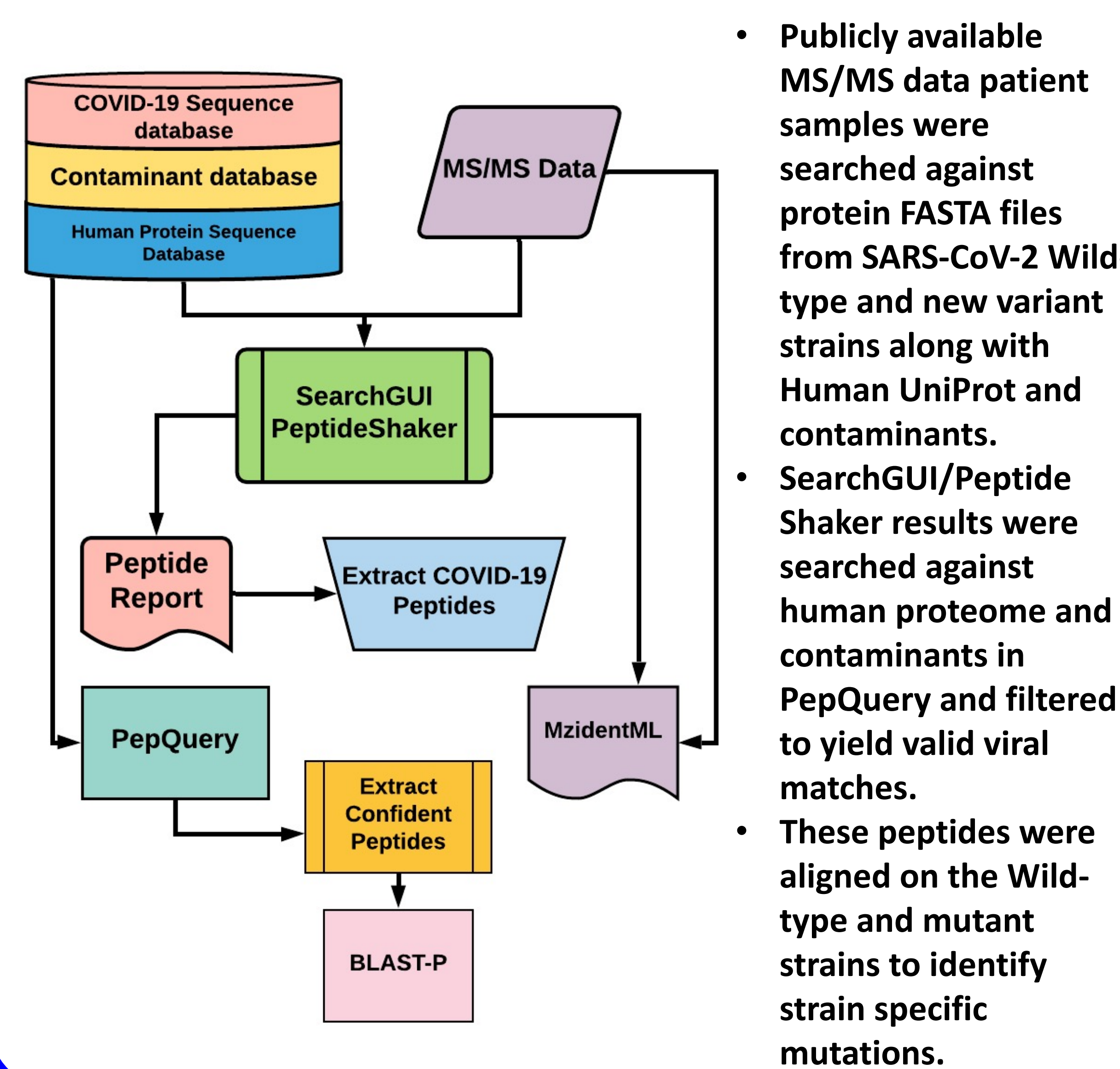
INTRODUCTION

- COVID-19 pandemic continues to wreak havoc across the world with the emergence of new strains.
- Due to emergence of these new strains, there is a need to advance the Mass Spectrometry-based COVID-19 diagnostics to identify optimal strain specific biomarker peptides from non-invasively collected clinical samples.
- The Galaxy-P team has developed workflows to detect and validate the SARS-CoV-2 peptides - optimal for targeted mass spectrometry-based biomarker assays.

EXPERIMENTAL METHODS

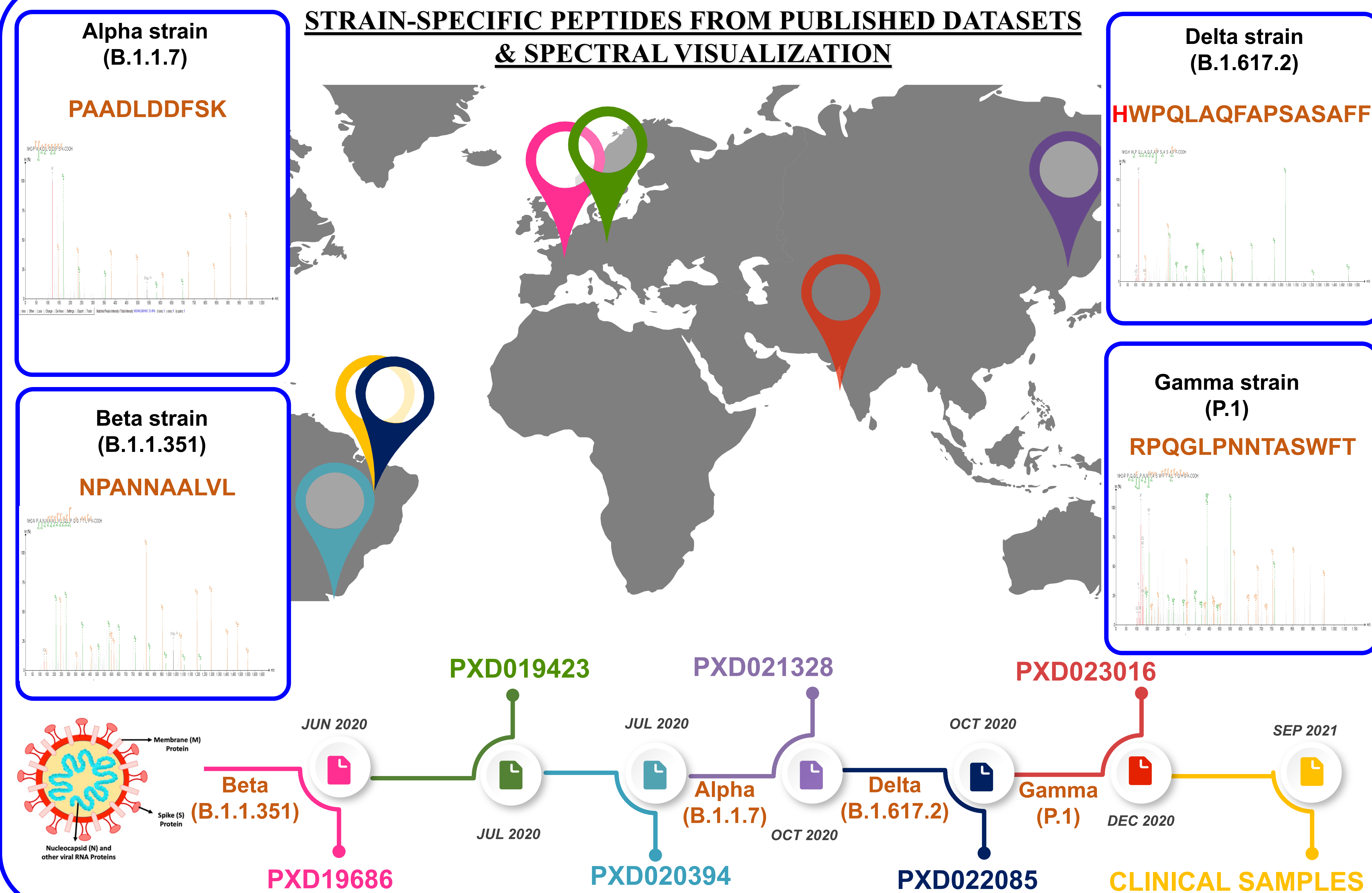
- Publicly available patient samples were searched against SARS CoV-2 proteome in a bioinformatics workflow in Galaxy (bottom).
- The discovery workflow provided us with few strain specific peptides.
- Our validation workflow was able to detect the presence of the peptides in these samples.

DISCOVERY WORKFLOW



- Publicly available MS/MS data patient samples were searched against protein FASTA files from SARS-CoV-2 Wild-type and new variant strains along with Human UniProt and contaminants.
- SearchGUI/Peptide Shaker results were searched against human proteome and contaminants in PepQuery and filtered to yield valid viral matches.
- These peptides were aligned on the Wild-type and mutant strains to identify strain specific mutations.

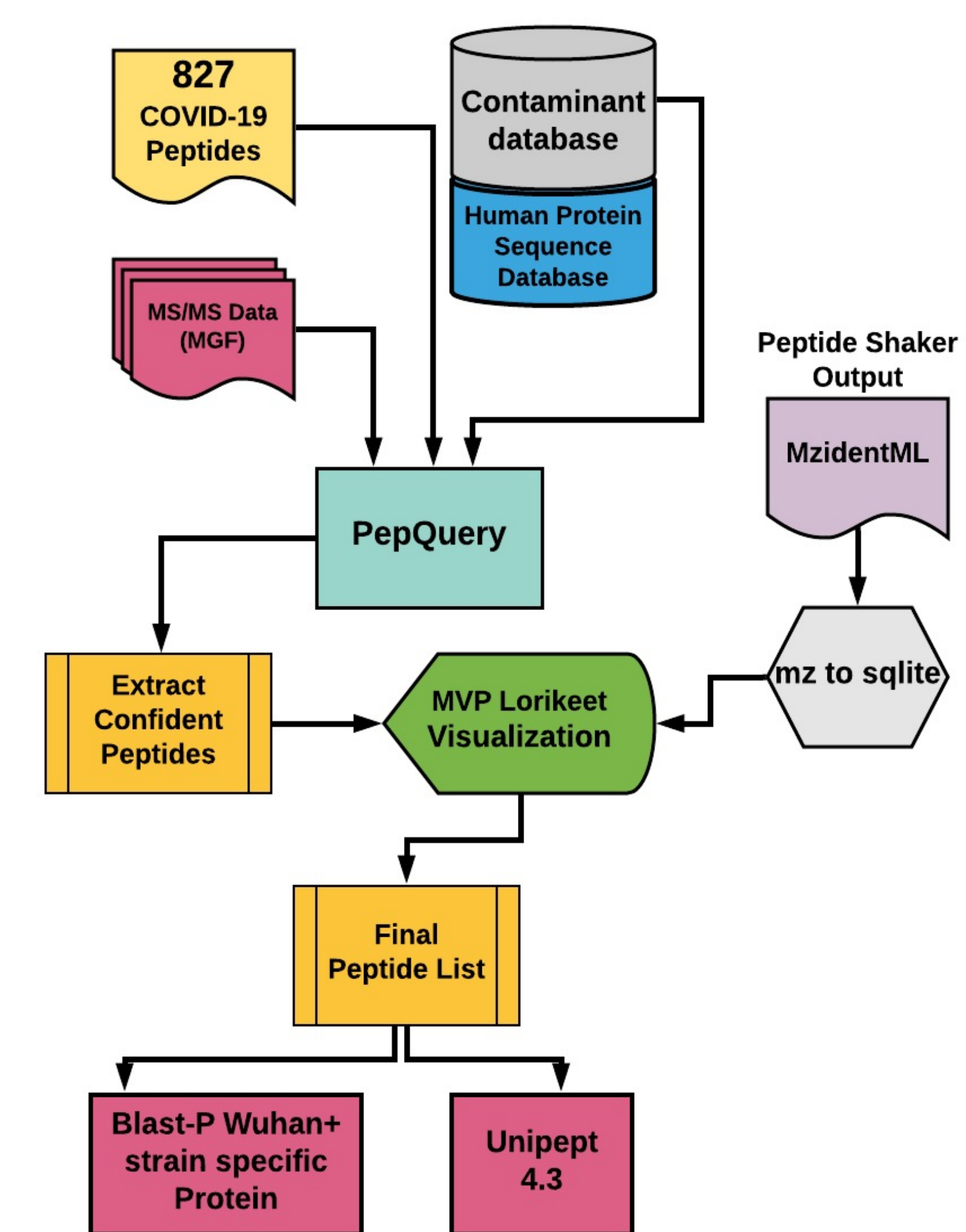
STRAIN-SPECIFIC PEPTIDES FROM PUBLISHED DATASETS & SPECTRAL VISUALIZATION



SPECIFICITY OF STRAIN-SPECIFIC PEPTIDES

PEPTIDE	PROTEIN	Pango Lineage	First observed	Last observed	First detected	Last detected	BLAST-P Identity (NCBI)	BLAST-P Identity (SARS-CoV-2 WT)	WHO NOMENCLATURE
GQGVPLNTSSPDDQLGY	{N}	{nan, B.1.1.395, D.4, B.1.1.397, B.1.404, B.1.1.243, B.1.1.242, B.1.1.241, B.1.1.240, B.1.1.244, B.1.1.249, B.1.549, B.1.548, B.1.1.394, B.1.541, B.1.540, B.1.1.71, B.1.1.70, B.1.1.72, B.1.1.75, B.1.1.74, B.1.1.77, B.1.1.77.9, B.1.1.77.8, B.1.1.49, B.1.1.45, B.1.1.47, B.1.1.46, B.1.1.40, B.1.1.43, B.1.1.42, A, U.1, U.2, U.3, B.1.1.141, B.1.1.142, B.1.1.416, AS.2, B.1.1.147, B.1.1.77.59, B.1.1.77.58, B.1.325, ...}	7/24/13	9/25/21	Bat	England	95.238	95.238	Delta, Alpha
NPANNAALVLQPLQGT	{N}	{B.1.2, AY.4, B.1.258.17, B.1.617.2, B.1.1.7}	1/4/21	9/14/21	USA	Germany	100	93.75	Alpha, Delta
LDDKPNFQDKVLLLNK	{N}	{B.1.617.2, AY.25, B.1.1.77, B.1.2, B.1, B.1.1.77.65, Q.8, B.1.621, B.1.1.77.58, AY.4, P.1, AY.7.1, B.1.36.8, B.1.1.528, B.1.1.363, B.1.1.7, B.1.1.348}	3/17/20	9/7/21	Germany	USA	94.118	94.118	Delta
LNTDHSSSDNLALLVQ	{M}	{B.1.1.515}	10/7/20	10/7/20	Germany	Germany	94.118	94.118	-
LGMEVTPSGTWLTY	{N}	{B.1, B.1.1.7}	1/25/21	2/10/21	Sweden	Sweden	94.44	94.44	Alpha
LTFGGPSDSTGSDQNGER	{N}	{P.1.12, B.1.1, B.1.1.77.53, AY.37, AY.30, B.1.438.1, B.1.360, B.1.429, P.1, B.1.427, P.2, B.1.617.2, AY.26, AY.5.2, AY.4, AY.5, AY.3, B.1.1.7, AY.9, B.1.1.389, B.1.258, A.5, B.1.2, C.37}	3/17/20	9/7/21	Spain	USA	94.44	94.44	Gamma, Delta
HWPQLAQFAPSASAFF	{N}	{B.1.617.2, B.1.351, P.1, B.1.1, B.1.1.420, B.1.1.7}	5/15/20	9/14/21	India	USA	100	93.75	Delta, Beta, Gamma
FTALTQH GK	{N}	{B.1.564}	8/24/20	8/28/20	USA	Brazil	100	88	-
SFNPETNLLNVLPHGTLT	{M}	{None}	4/7/20	6/21/21	Italy	France	88.89	88.89	-
RPLESELVGLAVLLR	{M}	{None}	4/7/20	6/21/21	Italy	France	88.89	88.89	-

VALIDATION WORKFLOW



- The PepQuery search method ranks the matches of our peptide panel to COVID19 MS data against a reference proteome containing proteins from human and contaminants.
- Resulting data was filtered to generate a list of confident peptide matches to SARS-CoV-2 and its variants.
- The peptide list was validated using bioinformatic tools such as BLAST-P and Unipept, and the spectral quality of these validated peptides were inspected using the Multi-omics Viewing Platform (MVP) or Proteomics data viewer (PDV) tools.

CONCLUSION

- Most of the strain-specific peptides, align to nucleocapsid and spike proteins of the viral particle and would serve as the optimal targets for direct detection of SARS-CoV-2 strains.
- The flexibility of the Galaxy workflows has the potential to detect the emerging strains on newly published MS datasets (DDA or parallel-reaction-monitoring (PRM)). Currently, we are expanding our peptide panel to include peptide targets to detect emerging strains.
- The Galaxy workflows along with the input data and updated results are available via <https://covid19.galaxyproject.org/proteomics/>.

ACKNOWLEDGEMENTS

We would like to acknowledge funding for this work from the National Cancer Institute - Informatics Technology for Cancer Research (NCI-ITCR grant 1U24CA199347) and National Science Foundation (grant 1458524.)

REFERENCES

1. Rajczewski, A.T., Mehta, S., Nguyen, D.D.A. et al. A rigorous evaluation of optimal peptide targets for MS-based clinical diagnostics of Coronavirus Disease 2019 (COVID-19). *Clin Proteom* 18, 15 (2021). <https://doi.org/10.1186/s12014-021-09321-1>
2. Peter S. Thuy-Boun, Mehta, S., Gruening, B., et al. Metaproteomics Analysis of SARS-CoV-2-Infected Patient Samples Reveals Presence of Potential Coinfecting Microorganisms. *Journal of Proteome Research* 2021 20 (2), 1451-1454. <https://doi.org/10.1021/acs.jproteome.0c00822>