


RESEARCH

Open Access



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 Nguyen¹, Björn Grüning³, James E. Johnson², Thomas McGowan², Timothy J. Griffin¹ and Pratik D. Jagtap^{1*} 

Abstract

Background: The Coronavirus Disease 2019 (COVID-19) global pandemic has had a profound, lasting impact on the world's population. A key aspect to providing care for those with COVID-19 and checking its further spread is early and accurate diagnosis of infection, which has been generally done via methods for amplifying and detecting viral RNA molecules. Detection and quantitation of peptides using targeted mass spectrometry-based strategies has been proposed as an alternative diagnostic tool due to direct detection of molecular indicators from non-invasively collected samples as well as the potential for high-throughput analysis in a clinical setting; many studies have revealed the presence of viral peptides within easily accessed patient samples. However, evidence suggests that some viral peptides could serve as better indicators of COVID-19 infection status than others, due to potential misidentification of peptides derived from human host proteins, poor spectral quality, high limits of detection etc.

Methods: In this study we have compiled a list of 636 peptides identified from Sudden Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) samples, including from in vitro and clinical sources. These datasets were rigorously analyzed using automated, Galaxy-based workflows containing tools such as PepQuery, BLAST-P, and the Multi-omic Visualization Platform as well as the open-source tools MetaTryp and Proteomics Data Viewer (PDV).

Results: Using PepQuery for confirming peptide spectrum matches, we were able to narrow down the 639-peptide possibilities to 87 peptides that were most robustly detected and specific to the SARS-CoV-2 virus. The specificity of these sequences to coronavirus taxa was confirmed using Unipept and BLAST-P. Through stringent p-value cutoff combined with manual verification of peptide spectrum match quality, 4 peptides derived from the nucleocapsid phosphoprotein and membrane protein were found to be most robustly detected across all cell culture and clinical samples, including those collected non-invasively.

Conclusion: We propose that these peptides would be of the most value for clinical proteomics applications seeking to detect COVID-19 from patient samples. We also contend that samples harvested from the upper respiratory tract and oral cavity have the highest potential for diagnosis of SARS-CoV-2 infection from easily collected patient samples using mass spectrometry-based proteomics assays.

*Correspondence: pjagtap@umn.edu

[†]Andrew T. Rajczewski and Subina Mehta contributed equally to this work

¹ Department of Biochemistry, Molecular and Cell Biology Building, University of Minnesota, 420 Washington Ave SE 7-129, Minneapolis, MN 55455, USA

Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords: Pandemic, Bioinformatics, Peptide-detection, Mass spectrometry, Viral proteome, Workflows

Introduction

In the latter half of 2019, a pneumonia-like disease arose in the Wuhan Province of China [1]. Subsequent analysis showed the cause to be a betacoronavirus initially called 2019-novel coronavirus (2019-nCoV). This disease soon spread throughout the world and came to be known as coronavirus disease 2019 (COVID-19) with the clinical classification Sudden Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). As of the writing of this manuscript, there are over 154 million patients infected world-wide with COVID-19, with a current global death toll sitting at over 3.2 million people [2]. Patients report a litany of symptoms, ranging from fever, cough, and muscle aches in mild cases to acute respiratory distress syndrome (ARDS), multiple-organ failure, and death in the most severe cases [3, 4].

While the development of therapeutic treatments for infected patients [5, 6] and the eventual development of vaccines against SARS-CoV-2 [7–9] are of great importance for the management of this disease, rapid and effective diagnosis of COVID-19 infection has been and continues to be of primary importance. Most testing strategies used in the diagnosis of active COVID-19 infections utilize quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR) of viral RNA in samples collected from patients [10, 11]. Rapid COVID-19 testing is generally performed on readily accessible patient-derived samples with high viral loads, such as nasopharyngeal swabs and saliva. To improve turnover time and increase the volume of tests that can be performed, innovations in RNA-based testing have been introduced to cut down on the time required. Testing protocols have been developed that eschew the isolation of RNA from patient samples, allowing for much faster RT-qPCR analyses [12]. In addition, techniques such as Reverse Transcription Loop-mediated isothermal AMplification (RT-LAMP) [13] and Specific High Sensitivity Enzymatic Reporter UnLOCKing (SHERLOCK) [14] diagnostics allow for rapid point-of-care detection of SARS-CoV-2 RNA without the need for sophisticated training in PCR.

While these techniques are generally fast and highly specific for viral RNA, improper sample collection, storage, or processing could result in the degradation of RNA yielding potential false negative tests. In addition, their reliance on sequence amplification using reverse transcriptases and DNA polymerases introduces the potential for false negatives through the inhibition of these enzymes by components of the sample [15, 16]. Due

to the better chemical stability of proteins compared to RNA, as well as the lack of a need for intermediary enzymes and signal amplification via PCR, clinical proteomics has emerged as a potential supplemental test for the diagnosis of COVID-19 through direct detection of viral peptides via LC-MS [17]. Specifically, targeted methods such as selected reaction monitoring (SRM) and parallel reaction monitoring (PRM) to detect peptides specific to the virus could be most useful in a clinical setting [18, 19]. However, not all the potential viral peptides derived from SARS-CoV-2 infection are equally suitable as targets, based on well-known limitations of targeted LC-MS methods for proteomics; some tryptic peptides of SARS-CoV-2 could have intrinsic physicochemical properties limiting their reproducible detection in a mass spectrometer, as well as co-elution from the LC with more abundant peptides that mask their presence in the sample. In addition, proteomics software can sometimes make putative peptide spectrum matches (PSMs) with spectra that are of poor quality, making for uncertain identification of peptides of interest [20, 21]. Additionally, a key requirement for targeting peptides for virus detection is that these are specific to the SARS-CoV-2 virus, with no potential overlap with other coronaviruses or other organisms.

In order to evaluate the most robustly detectable SARS-CoV-2 peptides, and make the detection of these viral peptides in human samples in a clinical setting all the more feasible, we set out to examine proteomic datasets from three cell culture-based studies [22–24] and seven clinical studies [25–30]. We utilized automated workflows implemented in the Galaxy platform and made accessible via the European Galaxy public instance to first identify as many SARS-CoV-2 peptides possible in all samples, creating a master list of SARS-CoV-2 peptides identified across the samples. We then interrogated these peptides using the PepQuery search engine [31] to confirm the quality of these PSMs and determine whether the matched sequences were unique to SARS-CoV-2 or could be better ascribed to the human proteome or that of another closely related coronavirus. Peptides and their associated PSMs which survived this rigorous filtering were then manually validated using the Multi-omics Visualization Platform [32] and further analyzed for specificity to the SARS-CoV-2 virus via BLAST-P [33] and MetaTryp [34]. Taken together, our analyses enable the construction of a high-confidence target peptide list that would form the basis of a targeted clinical proteomics assay for SARS-CoV-2 infection.

Methods

Case study

For establishing workflows to evaluate virus-specific peptides, three published cell culture datasets [22–24] which used SARS-CoV-2 infected Vero cell lines were chosen, along with five clinical datasets [26–29, 35].

Cell culture datasets

Gouveia et al. published a dataset (PXD018804) with SARS-CoV-2 infected Vero cells from *Chlorocebus* primates to generate a high-resolution mass spectrometry dataset. The second dataset was published by Grenga et al. (PXD018594) wherein a seven-day time course shotgun proteomics study was performed on Vero E6 cells infected by Italy-INMI1 SARS-CoV-2 virus at two multiplicities of infection. The third cell culture dataset chosen was published by Davidson et al. (PXD018241), which also utilized Vero E6 cells to investigate the viral transcriptome and proteome.

Clinical datasets

The first clinical dataset chosen was from the study by Cardozo et al. (PXD021328), wherein they collected bottom-up mass spectrometry (MS) data on combined oropharyngeal and nasopharyngeal samples from ten COVID-19 positive patient samples. A second clinical dataset was from the Ihling group (PXD019423) to detect SARS-CoV-2 virus proteins from saline gargle samples of COVID-19 infected patients. The third dataset was obtained from the Rivera group (PXD020394) comparative quantitative proteomic analysis from oro- and naso-pharyngeal swabs used for COVID-19 diagnosis was performed. Further, unanalyzed oro/nasopharyngeal data from Cardozo et al. [25] (PXD025214) as well as a nasopharyngeal swab dataset from Bankar et al. [30] (PXD023016) were interrogated for the presence of our proposed targets. Datasets derived from COVID-19 patient lung biopsies (PXD018094) and bronchoalveolar lavage fluid (BALF) (PXD022085) were analyzed to determine the utility of our workflow to identify SARS-CoV-2 in clinically relevant sample types.

Sequence database searching

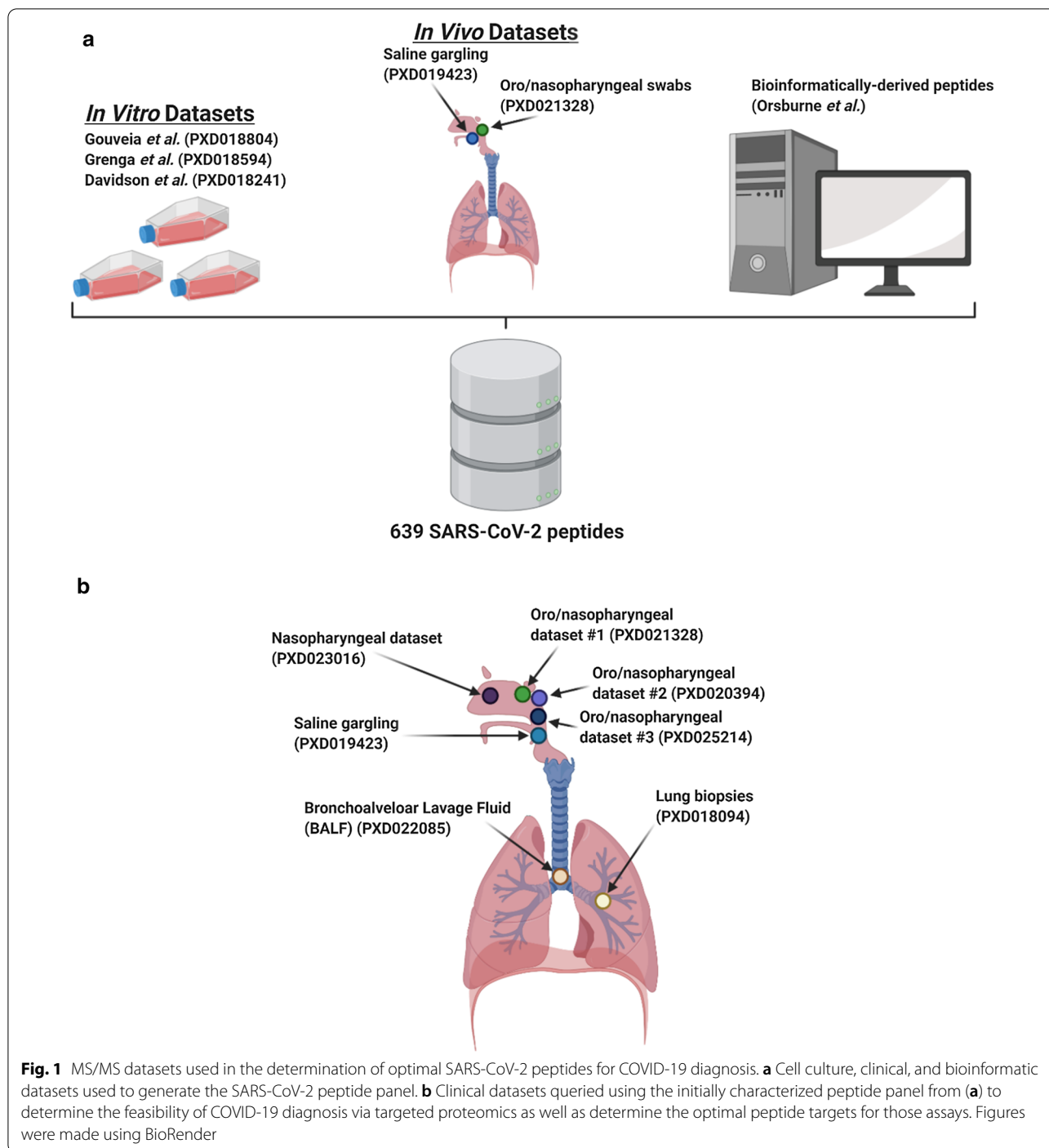
The Galaxy workflow for peptide identification (Figs. 1, 2a) includes conversion of RAW data to MGF and mzML format. In case of the cell culture study, the MGF files are searched against the combined database of *Chlorocebus* sequences, contaminant proteins (cRAP) and SARS-CoV-2 proteins. For the clinical database, the resultant MGF files were searched against the combined database

of Human Uniprot proteome, contaminants, and SARS-CoV-2 proteins database.

For sequence database searching in the workflow, search algorithms—X! tandem, MSGF+, OMSSA were used within SearchGUI [36] to produce PSMS, followed by False Discovery Rate (FDR) and protein grouping analysis using PeptideShaker [37]. The search parameters for digestion, modifications, tolerance, and FDR were chosen accordingly from the published papers for each of these datasets (Additional file 1: Data S1). The peptide report generated using PeptideShaker was used to extract confident COVID-19 peptides. The peptides were validated using PepQuery analysis with MS tolerance of 10 ppm and MS/MS tolerance of 0.05 Da. The SARS-CoV-2 peptides detected from the three cell culture datasets and two clinical datasets were merged with peptide list from in silico analysis of genomic sequences by Orsburn et al. [38] to generate a peptide panel for interrogation of clinical data sets. The re-analysis of the dataset using the workflow is available online on the COVID-Galaxy website (<https://COVID19.galaxyproject.org/proteomics>) and the workflows and outputs can be found online (see Data and Workflow Availability).

Peptide validation

This SARS-CoV-2 peptide panel was subjected to the Peptide Validation workflow (Fig. 2b) against the clinical datasets specified above. The peptide validation workflow includes re-analysis by PepQuery as well as manual visualization and inspection in the Lorikeet application of Multi-omics Visualization Platform (MVP) to ascertain the quality of peptide sequences matched to MS/MS spectra. Unrestricted modification searching and amino acid substitutions were enabled in PepQuery to ensure the most rigorous search possible, with hypothetical post-translational modifications and amino acid substitutions applied to the reference peptides to examine every possible sequence match to the putative SARS-CoV-2 spectra. To rule out misidentification of host peptides and ensure the specificity of validated peptides for the SARS-CoV-2 virus, a reference proteome of human proteins as well as the proteomes of SARS-CoV, OC43, NL62, HKU1, 229E, SARS-MA15, SARS-WIV1, and MERS-CoV were used for this rigorous evaluation. The results from PepQuery were then filtered to remove any peptides which had matches to the reference proteomes, leaving only those peptides which aligned to the SARS-CoV-2 proteome. The spectra of the validated peptides were then manually annotated using the Multi-omics Visualization Platform (MVP) [32] or the Proteomics Data Viewer (PDV) [39] to ensure the quality of the potential SARS-CoV-2 targets. The workflow also included additional,

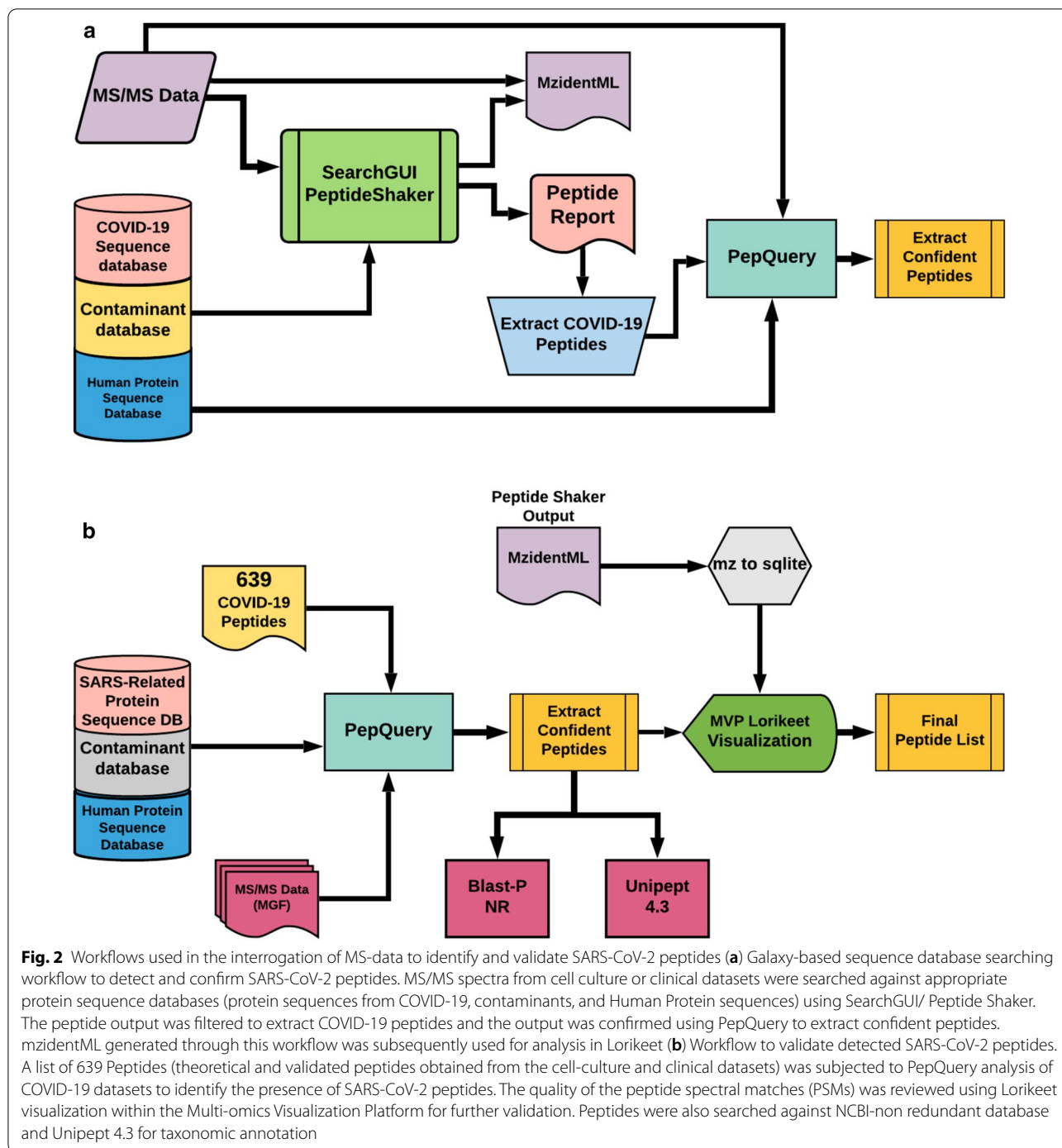


optional in-line characterization of these peptides by searching against NCBI-non redundant (nr) BLAST-P and Unipept [40] analysis. Further offline analysis was performed using NCBI BLAST-P analysis as well as the MetaTRYP [34] coronavirus database. The peptide validation workflow can be found at COVID Galaxy website (<https://COVID19.galaxyproject.org/proteomics>).

Results

Sequence database searching results

Sequence database searching to generate peptide spectral matches (PSMs) and identify peptides from three cell culture datasets (Fig. 1a) using the workflow shown in Fig. 1a led to detection of 139 peptides, 99 peptides and 579 peptides, respectively. For the two clinical datasets



analyzed using the workflow, we detected 76 and 8 peptides, respectively (Table 1). These peptides together represented 630 unique peptides corresponding to several proteins coded in the SARS-CoV-2 genome; to these we then added a further 9 unique peptides generated from in silico translated data by Orsburn et al. [38] to generate a list of 639 unique SARS CoV-2 peptides (Additional

file 1: Table S1). This 639-peptide panel was further used to interrogate the clinical datasets and determine the reliability of their detection using un-targeted MS-based proteomics. BLAST-P analysis of the 639-peptide panel showed that these peptides mapped to 27 proteins and open reading frames within the SARS-CoV-2 genome (Fig. 3), with sequence coverage ranging from

Table 1 Peptides generated from MS datasets

	Manuscript (Proteome Xchange ID)	SARS-CoV-2 peptides detected using Database Search Workflow	Detected peptides	SARS-CoV-2 peptides detected using Peptide Validation Workflow	
Cell-culture datasets	<i>Gouveia</i> et al. (PXD018804)	139 peptides	630 distinct peptides	–	
	<i>Grenga</i> et al. (PXD018594)	99 peptides		–	
	<i>Davidson</i> et al. (PXD018241)	579 peptides		–	
Clinical datasets	<i>Cardozo</i> et al. (PXD021328)	76 peptides		70 peptides	87 distinct peptides
	<i>Ihling</i> et al. (PXD019423)	8 peptides		21 peptides	
	<i>Rivera</i> et al. (PXD020394)	–	–	10 peptides	
	<i>Leng</i> et al. (PXD018094)	–	–	14 peptides	
	<i>Zeng</i> et al. (PXD022085)	–	–	37 peptides	
	<i>Cardozo</i> et al. (PXD025214)	–	–	39 peptides	
	<i>Bankar</i> et al. (PXD023016)	–	–	35 Peptides	

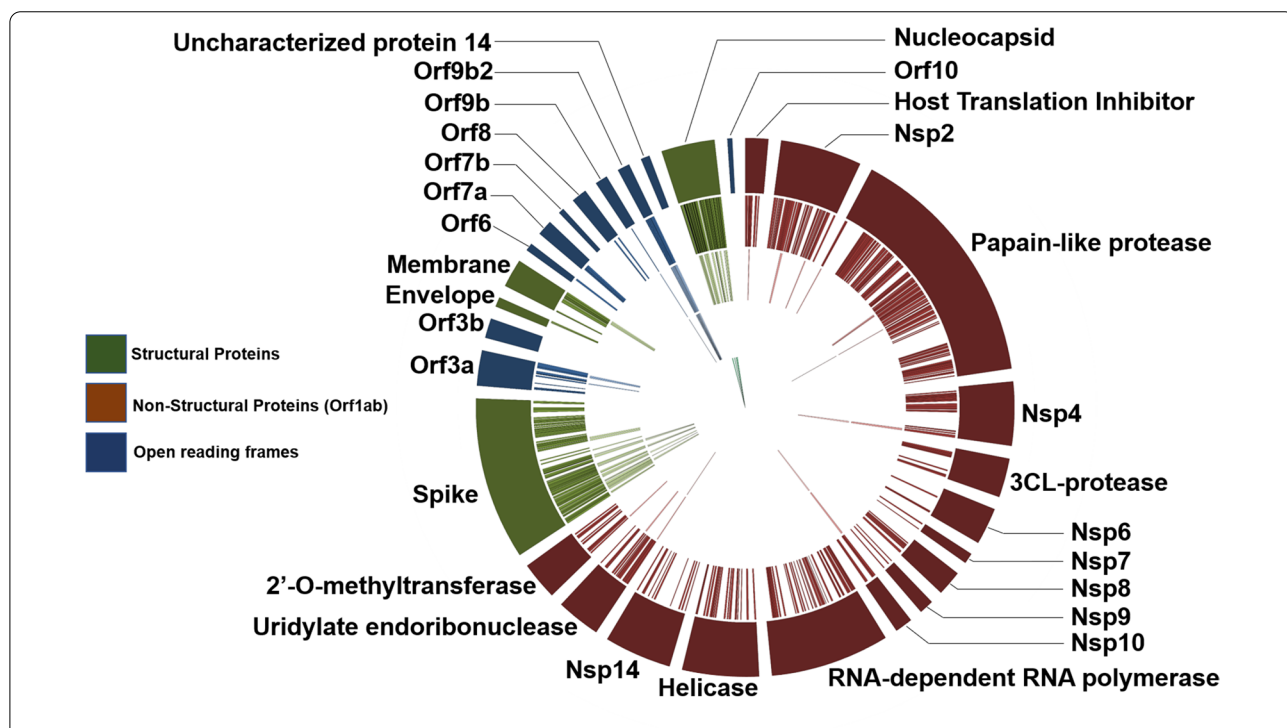


Fig. 3 Protein assignment of detected and validated SARS-CoV-2 peptides: Circos plot of peptides against SARS-CoV-2 proteins (outermost ring). Of the 639-peptide panel (2nd outermost ring), many peptides could be identified using our validation workflow in clinical and cell culture datasets (3rd outermost ring). Peptides derived from ORF9b, papain-like protease, Nsp4, Nsp10, uridylate endoribonuclease (Nsp15) and certain spike protein peptides were only found in cell culture datasets (2nd innermost ring). Final peptides chosen for targeted analysis are annotated in the innermost ring. Circos plot was generated in Galaxy [65]

4.7% coverage (Proofreading exoribonuclease Guanine-N7 methyltransferase protein) to 93.7% coverage (Nucleocapsid protein) (Additional file 1: Figure S1).

Peptide validation results

Having derived a comprehensive panel of 639 peptides detected across multiple COVID-19 datasets, we then

utilized a validation workflow based around the PepQuery database to interrogate the dataset PXD020394, derived from oro- and naso-pharyngeal swabs collected in the clinic from patients positive and negative for COVID-19. This resulted in detection of 10 SARS-CoV-2 peptides from our panel in these clinically relevant samples (Additional file 1: Figure S2).

We detected eight of the peptides in COVID-19 positive sample replicates—with the peptide RGPEQTQGN-FGDQELIR being detected in all positive sample replicates, followed by TATKAYNVTQAFGR and AYN-VTQAFGR detected in 6 out of 10 replicate samples (Additional file 1: Figure S2). We also detected two peptides— GVEAVMYMGTLSEYQFK and CDLQNYGDS-ATLTPK— from COVID-19 negative samples.

We also re-analyzed the clinical datasets used in the generation of the 639 panel (the second oro/nasopharyngeal dataset from Cardozo et al. as well as the saline gargling dataset), using our validation workflow. The validation workflow provides a complementary method to the initial sequence database searching method for confirming peptide spectrum matches, based primarily on the PepQuery tool. For the oro/nasopharyngeal dataset, we confirmed confident identification of 70 peptides using the peptide validation workflow (as compared to 76 detected using the initial sequence database searching workflow). For the saline gargling dataset, we confirmed the presence of 21 peptides using the peptide validation workflow (as compared to 8 peptides detected using the peptide search workflow). Considering all peptides detected in clinical samples using the peptide validation workflow, we detected 87 peptides with confidence (Table 1). These validated peptides were assigned to known proteins from the COVID-19 proteome. Most of the peptides detected in the upper respiratory tract were aligned to structural proteins making up the viral capsid such as nucleocapsid protein N, the viral matrix protein M, and the spike protein S; fewer peptides were aligned to proteins involved in viral replication such as papain-like protease, RNA-directed RNA polymerase, non-structural protein, 2'-O-methyltransferase and host translation inhibitor (Fig. 3). The most peptides were identified in the oro/nasopharyngeal dataset that consisted of combined oropharyngeal and nasopharyngeal swabs analyzed by Cardozo et al.; fewer peptides were identified from PXD019423 and PXD020493, which were derived from gargled saline samples and a second study of combined oropharyngeal and nasopharyngeal samples, respectively.

Based on the sample-type from which they were detected (clinical samples versus in vitro cell culture experiments) and their source (empirically derived from MS/MS data versus theoretically determined based on genomic sequence data), we categorized them as being present or absent in the various datasets based on their confident detection using our validation workflow. We found that the validated peptides clustered into distinct groups based on their source sample and dataset of origin, and how they were originally identified (Additional file 1: Table S1). Eleven peptides were found to be highly

consistent across the upper respiratory clinical datasets as well as the in vitro cell culture datasets. In considering theoretical peptides proposed by the Orsburn et al., eleven of those predicted peptides were in clinical samples and eight were detected in the in vitro cell culture samples. Twenty-two SARS-CoV-2 peptides that were not initially identified using the database search workflow were identified by matching to MS/MS spectra using the PepQuery-based validation workflow across multiple datasets.

Having established the presence of validated SARS-CoV-2 peptides in our initial clinical datasets, we then interrogated additional clinical datasets to further validate the utility of our methodology. Further patient datasets comprising oro/nasopharyngeal swabs (PXD025214) as well as nasopharyngeal datasets from COVID-19-positive patients (PXD023016) were analyzed using the PepQuery validation workflow and the 639-peptide panel. Analyses of these datasets revealed 39 and 35 validated peptides, respectively, which had considerable overlap with our initial analyses of oro/nasopharyngeal and gargling datasets. Clinical datasets from lung biopsies (PXD018094) and BALF (PXD022085) were also interrogated to determine the applicability of our approach in detecting SARS-CoV-2 within the deeper respiratory tract. Our validation workflow was able to confidently match MS/MS to 15 peptides in the lung biopsy dataset and 37 peptides in the BALF dataset. In comparing the peptides found within the upper respiratory samples to those detected within the lung biopsy samples and the BALF samples, the majority of the peptides detected in the deep lung datasets are unique to the samples being analyzed, with no peptides in common with the upper respiratory tract samples (Additional file 1: Table S1). Despite this apparent disparity, BLAST-P analysis reveals the alignment of SARS-CoV-2 peptides identified in deep lung tissue corresponding to a similar complement of SARS-CoV-2 proteins as the upper respiratory tract datasets, including additional structural proteins such as the Spike protein and Membrane glycoprotein as well as other nonstructural and replication proteins such as RNA-directed RNA polymerase, Protease 3CL-PRO, etc. In addition, the lung biopsy and BALF datasets also included MS-data from patients negative for COVID-19. In contrast to the two SARS-CoV-2 PSMs identified in the oro/nasopharyngeal samples from COVID-19-negative patients, samples analyzed from lung biopsies of COVID-19-negative patients resulted in identification of 21 SARS-CoV-2 peptides. Similarly, 37 peptides were detected in BALF samples isolated from patients that tested negative for COVID-19.

The last category of peptides that we evaluated were detected from COVID-19 cell culture studies (Additional

file 1: Table S1, Figure S3). These peptides were derived from protein sequences that were not available in the initial Uniprot sequence databases but were subsequently added as more COVID19 strains were sequenced [41, 42]. We added these sequences to the sequence database to enable the detection of these COVID-19 proteoforms. Using this updated sequence database, we detected and validated twelve peptides from Accessory protein ORF9b from SARS-CoV-2 and two peptides from ORF1ab polyprotein from SARS-CoV-2. These peptides were observed only in the cell culture datasets, and not in the clinical datasets (Fig. 3).

Identifying detected peptides with highest spectra quality.

As a quality check on our bioinformatic workflows, we utilized the Multi-Omics Visualization Platform [32] and Proteomics Data Viewer to manually assess the spectral quality of the peptides that passed PepQuery validation, as well as elucidate the distribution of these peptides throughout the six datasets we analyzed. It is critical that the peptides used for targeted MS-based assays for detecting SARS-CoV-2as targets have excellent spectral quality to ensure adequate reliability in detecting and quantifying these peptides across a variety of clinical samples. Here, we focused on four peptides (AYNVTQAFGR, MAGNGGDAALALLLDR, RGPEQTQGNFGDQELIR, DGIWVATEGALNTPK) found in the positive patients from the second oro/nasopharyngeal dataset (PXD020934) that were also seen in the other clinical datasets as well as two peptides found in the negative patients (CDLQNYGDSATLPK,

GVEAVMYMGTLSEYQFK) from the same oro/nasopharyngeal dataset as benchmark examples for manually validating our spectra. For these selected four peptides, from the virus-positive samples we found largely complete b- and/or y-ion series with at least three consecutive ions detected in either series (Additional file 1: Figure S3). In addition, we found that these fragment MS2 ions showed intensities at least three-fold higher than the background noise level of the spectra. By contrast, the two peptides found in the negative samples had a very few fragment MS2 ions detected which scarcely rose above the level of the background noise (Additional file 1: Figure S3). Together, the MS/MS spectra of these six peptides were used to generate guidelines which were then used to manually interrogate the rest of the SARS-CoV-2 spectra as being genuine or misidentified by the bioinformatics software (Fig. 4). Manual annotation of the MS/MS spectra found that 16 of the peptides validated in PepQuery had MS/MS spectra suitable for confident identification.

As a part of our investigation, we detected and validated eight peptides that were predicted by Orsburn et al. [38] (Additional file 1: Table S1, Figure S3). However, Lorikeet visualization of the Peptide Spectral Match (PSM) quality detected only two peptides (with sequences ADETQALPQR and FDNVLPFNDGVY-FASTEK) in the clinical sample PXD021328 dataset; of these the ADETQALPQR was also detected in all three cell cultures sample datasets while the FDNVLPFNDGVYFASTEK sequence peptide was detected in two of the three cell culture samples (Additional file 1: Table S1,

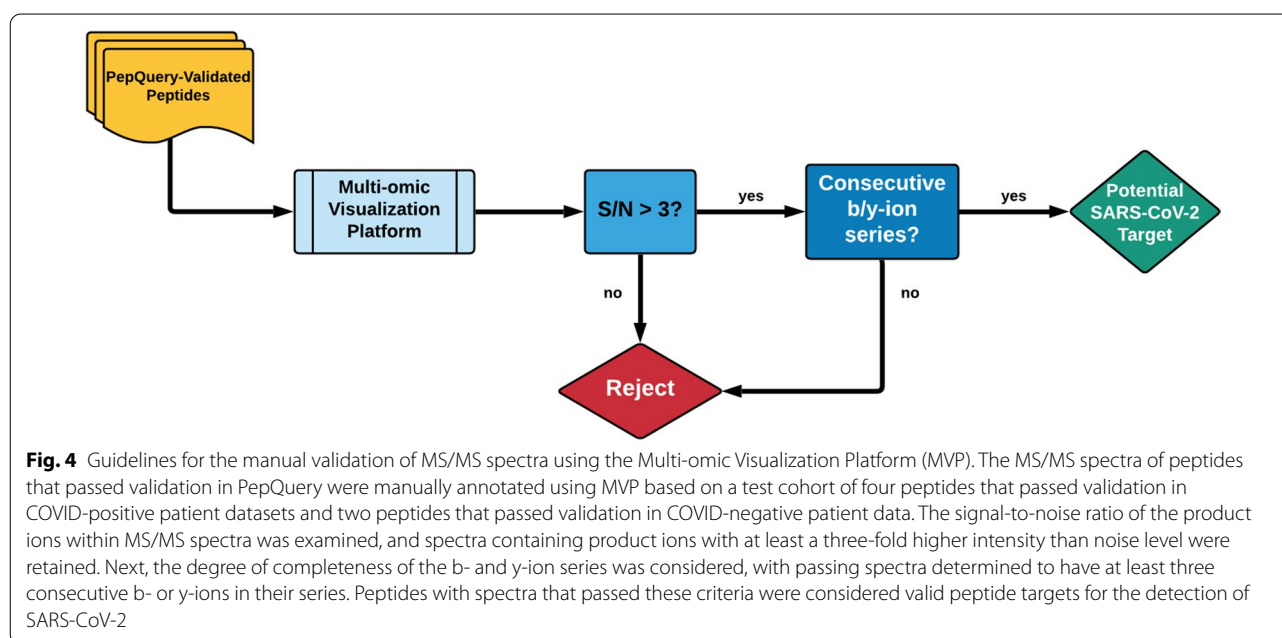


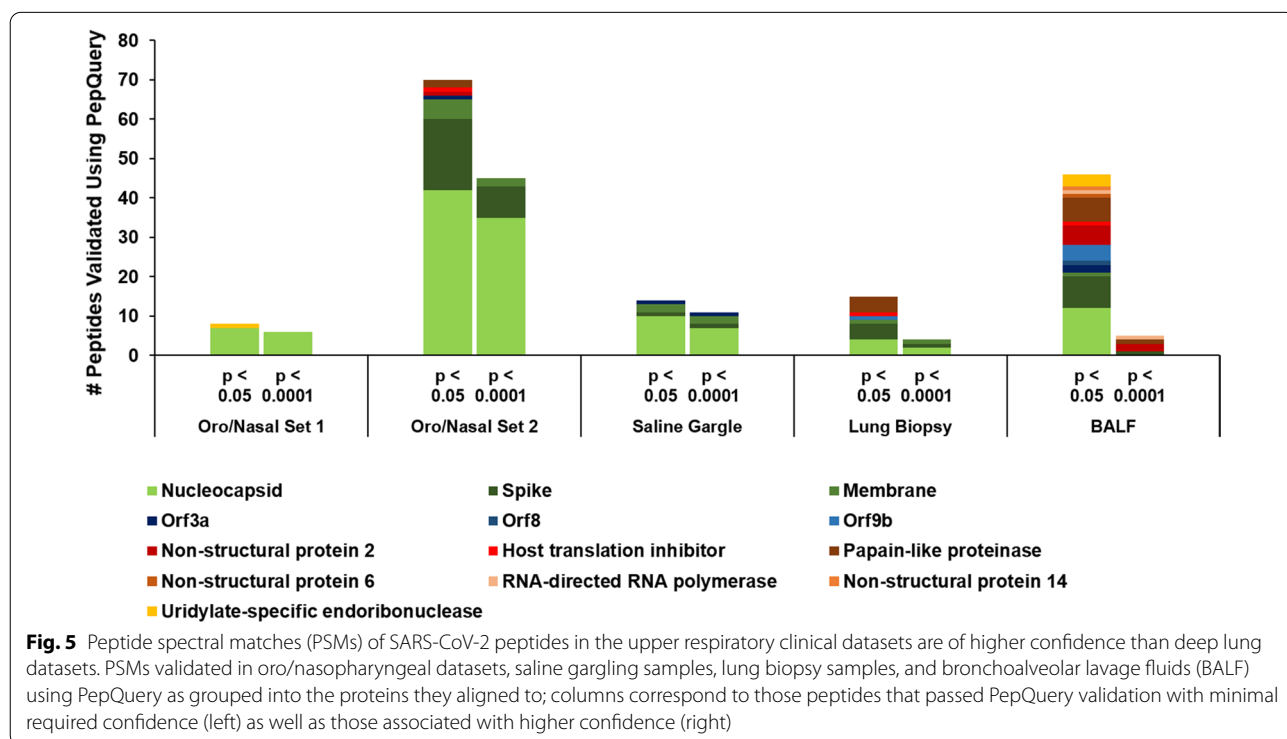
Figure S3). All the eight peptides were found to have good quality of PSMs in the cell culture datasets by using manual validation. Out of these eight peptides, a peptide with sequence HTPINLVR was detected in all cell culture experimental datasets (Additional file 1: Table S1).

We were able to validate 22 peptides using PepQuery which were not detected in the database search workflow (Additional file 1: Table S1). Subsequent manual validation of these peptides determined only two peptides had good quality spectra. The peptide of sequence DGIIWVATEGALNTPKDHIGTR was validated by using PepQuery and manual visualization in the PXD019423 dataset along with another peptide with sequence FTALTQHGKEDLK from the PXD02132 dataset (Additional file 1: Figure S3).

To determine the optimal candidates for the detection of SARS-CoV-2 using clinical MS-based assays, we resolved to focus on those peptides that passed PepQuery with the highest confidence, and subject these to manual inspection of spectral quality. We therefore sorted the results of our PepQuery analyses to include only those which had the highest confidence possible (p -value < 0.0001) to maximize the likelihood of passing our spectral annotation thresholds. In filtering the clinical datasets, we see a notable difference between the datasets derived from the upper respiratory tract (oro/nasopharyngeal datasets 1 and 2 as well as the saline gargling dataset) and those derived from deep lung tissue

(the lung biopsy and BALF datasets) (Fig. 5). In filtering the PepQuery results from the upper respiratory tract datasets, we noted that the structural proteins that had the most identified peptides- the nucleocapsid, membrane protein, and spike proteins- show relatively little elimination of PSMs, while the proteins involved in viral replication are generally lost, indicating relatively high confidence in the PepQuery validation of the peptides of the viral structural proteins. By contrast, peptides found in all proteins in the lung biopsy and BALF datasets were filtered out at this step, yielding only 3 and 4 high-confidence peptides in each dataset, respectively, leaving single peptides of nucleocapsid, membrane protein, and spike protein in the lung biopsy samples and single peptides of the spike protein, papain-like protease, non-structural protein 2, and RNA-dependent RNA polymerase.

The spectra of those peptides found to have high confidence in the clinical datasets were then analyzed using MVP, which leverages the Lorikeet viewer for visualization of annotated peptide MS/MS spectra. Manual analysis of the high-confidence peptides detected in the lung biopsy and BALF datasets using our previously established guidelines showed only the single peptide FLALCADSIIIGGAK, a component of Non-structural protein 2, in the BALF dataset as having a good quality spectrum, suggesting that the use of clinical samples collected using more invasive methods from deep within the lung



may be unsuitable for detection of SARS-CoV-2 using a clinical proteomics strategy. In contrast, 11 peptides in the upper respiratory tract datasets had high confidence and high-quality MS/MS-spectra. Of these, we then chose four peptides- MAGNGGDAALALLLDR, DGI-IWVATEGALNTPK, RGPEQTQGNFGDQELIR, and IGMEVTPSGTWLTYTGAIK, which were each, identified in at least three of the five upper respiratory clinical datasets, determining these to be the most reliable peptides for proteomics-based detection of SARS-CoV-2 in clinical samples harvested from the upper respiratory tract (Fig. 6, Additional file 1: Table S2). We assert that these represent the best candidates for targeted proteomics screening for potential cases of COVID-19.

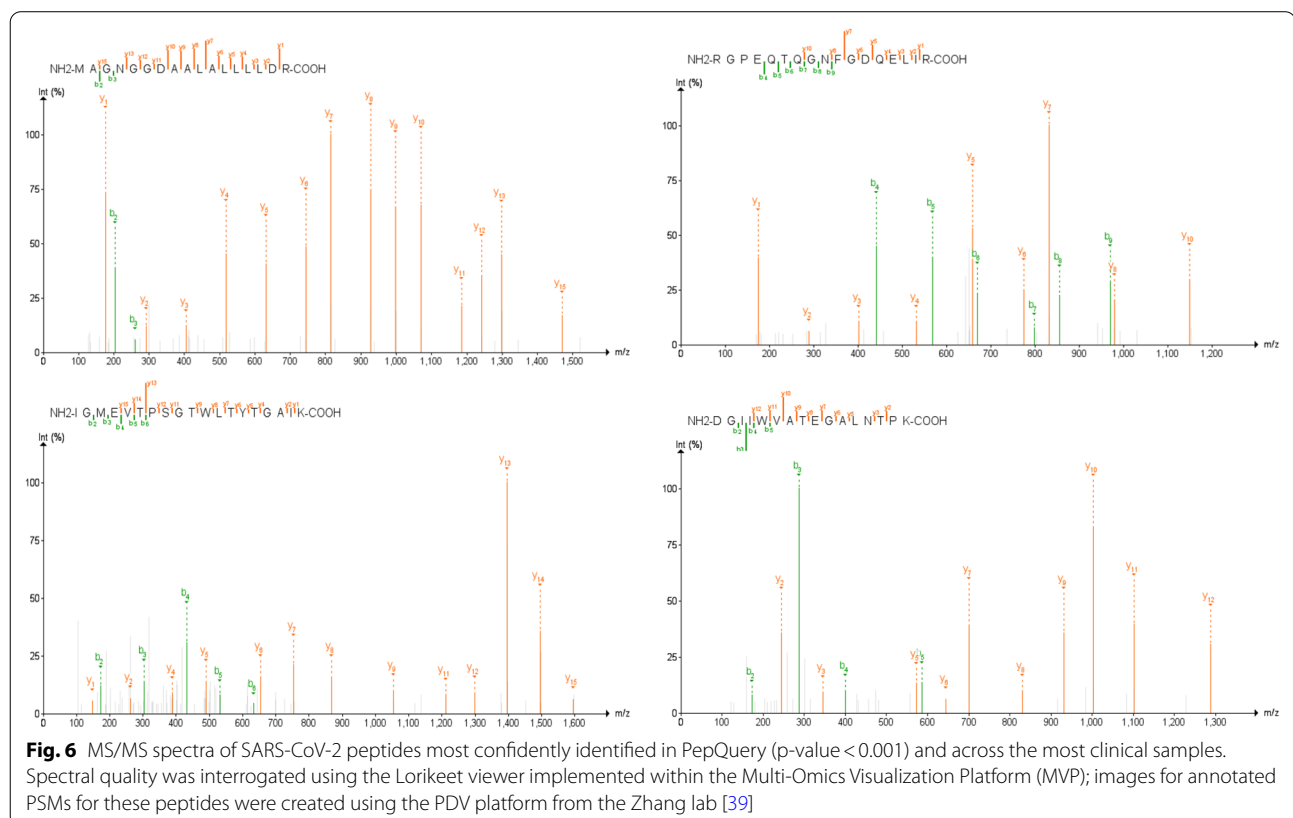
Viral specificity of high-quality peptides detected in SARS-CoV-2

We performed taxonomic analysis using MetaTryp to validate the specificity of the four highest-quality peptides detected in clinical samples to coronaviruses (Fig. 7a). Using this we found that these peptides mapped to proteomes of several coronaviruses, with each showing alignment SARS-CoV-2. To gauge the degree of specificity of these peptides for SARS-CoV-2 over other coronaviruses and their potential human host, we performed

BLAST-P analysis of these peptides against proteomes for SARS-CoV-2, humans, and eight known pathogenic human coronaviruses. To interrogate all possible matches to the target organisms, a relatively lax E-value cutoff of 1 was used. In considering the sequence alignment of these peptides, the peptides examined found a high degree of alignment to the nucleocapsid protein (N-protein) of SARS-CoV-2 (Fig. 7b). Each of the four distinct peptides that showed alignment to the N-protein also showed 100% sequence homology uniquely to SARS-CoV-2, with decreased sequence alignment in other closely related coronaviruses. One peptide sequence, MAGNGGDAALALLLDR, showed perfect alignment to the SARS-CoV-2 nucleocapsid protein with no alignment to the same protein in any other viruses. In all cases, no alignment to any human proteins was noted.

Discussion and conclusions

Clinical diagnostics using targeted MS-based proteomics has found considerable utility in recent years as a powerful tool for detecting peptide biomarkers characteristic of several diseases. Bottom-up proteomics has been used to characterize tumors in biopsied breast cancer tissues [43, 44], to explore the phenotypic changes that occur with opportunistic fungal infections in HIV/AIDS



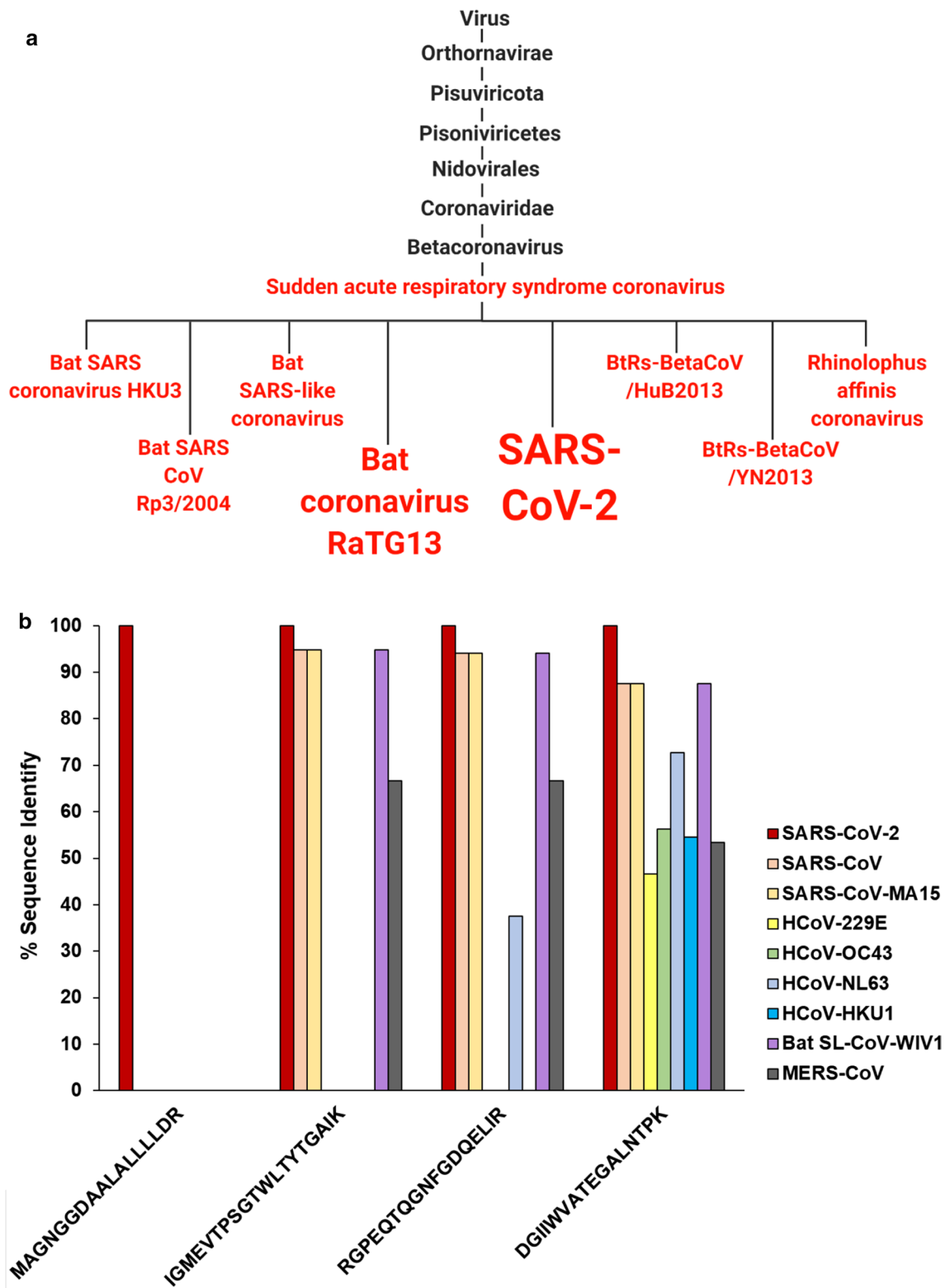


Fig. 7 Specificity of target peptides as for coronaviruses and for SARS-CoV-2 (a) MetaTryp taxonomic analysis of the 4 most consistently found peptides. Coronaviruses with matches to peptides are highlighted in red and font size is correlated with the number of peptides that show a match in that coronavirus. Created with BioRender.com (b) Sequence identity of peptides that show BLAST-P alignment with viral nucleocapsid protein

patients [45], and even differentiate between COVID-19 patients at differing WHO severity grades [46]. While these experiments effectively measure the phenotype of patients to infer a disease state, direct detection of proteins using targeted MS-based methods (SRM) from disease organisms can be used as a diagnostic assay for diseases. For these, it is critical that the most reliable peptides, specific to the protein of interest, are determined.

The pressing nature of the COVID-19 pandemic presents an opportunity for the use of targeted MS-based proteomics to supplement conventional RT-qPCR diagnostic procedures [11] to mitigate the false negatives inherent in the detection of viral RNA [47], along with other advantages of direct detection of peptides, such as chemical stability of the target molecules. Ideally, direct detection of diagnostic peptides would be achieved in samples easily collected in the clinic using non-invasive methods. While many labs have begun proteomic analysis of samples to identify SARS-CoV-2 infection in both in-vitro models and clinical samples, the development of targeted assays based on this work requires preliminary work to determine those peptides which are most reliably detected and most specific for unambiguous diagnosis of infection. To mitigate this and establish the best targets possible for a SARS-CoV-2 clinical proteomics assay, we identified detectable SARS-CoV-2 peptides using Galaxy-based workflows. To narrow this list down to the most confident and reliably detected peptides, we then utilized a bioinformatics workflow built around the PepQuery search engine. Developed by Wen et al. [31], this search engine interrogates raw mass spectrometry data for spectral matches to pre-chosen peptide sequences of interest and compares these matched spectra to reference proteomes to see whether the peptide of interest is a better match to the data than any reference peptide, scoring the peptide match much faster and with much less processing power needed than a conventional sequence database search. By using PepQuery on peptides that have already been designated as potential matches, we can utilize the increased statistical power of using multiple peptide search engines [48] common to many proteomics software suites on a much faster time scale. Using this as well as other tools available in the Galaxy platform we were able to interrogate publicly available data to ascertain the most reliable peptides for detecting SARS-CoV-2.

In the two oral/nasopharyngeal datasets and gargled saline dataset we examined, we found 75 peptides within the original list of 639 detected peptides that showed a high-confidence match to SARS-CoV-2 proteins over human proteins or other coronavirus proteins, suggesting that the unambiguous detection of SARS-CoV-2 in patients using proteomics technology is theoretically possible. These peptides were found in proteins

throughout the viral particle (Fig. 3), with more structural protein peptides detected than replication proteins. It was observed that the datasets stemming from the clinical samples had noticeably fewer peptides validated in them compared to those from in vitro experiments; this is potentially due to larger amounts of material, the differential abundance of host proteins in clinical samples compared with cultured samples [49], and the lack of viral clearance from cultured cells [50]. Of these, manual annotation found that 16 peptides could be truly said to have good quality MS/MS spectra, based on our thresholds for PSM quality and annotation.

From the 16 validated peptides with high-quality spectra, 11 peptides also were known to be high confidence matches in PepQuery. From these we chose four peptides that had high-confidence matches in PepQuery, were consistently seen in clinical samples, and were unique to SARS-CoV-2, making them the best candidates for diagnosis of COVID-19 using targeted MS-based methods. Given their high degree of specificity for SARS-CoV-2 and the high quality of their spectra, we postulate that the detection of any of these individual peptides in a clinical patient would warrant further clinical investigation of the patient's infection status. It is notable that these are all found within the nucleocapsid phosphoprotein, or N-protein. The nucleocapsid phosphoprotein is common to coronaviruses and serves to complex with and stabilize the viral RNA genome and package it into the viral particle [51, 52]. The viral ribonucleoprotein complex of N-protein and gRNA is localized beneath the matrix proteins (M-proteins) and spike proteins (S-proteins) that make up the capsid surface [53, 54]. As many copies of N-protein are needed to stabilize the viral gRNA, the N-protein is thought to be one of the most abundant proteins in the assembled SARS-CoV-2 viral particle [55]; analysis of SARS-CoV transcript levels in infected cells show the N-protein to be the most abundant RNA-based sub-genome within the cell [56]. Taken together, these phenomena explain the prominence of N-protein peptides across the proteomic datasets we examined. As the N-protein is a frequent amplification target for RT-qPCR assays as per FDA guidelines for diagnosis [57], we believe that our results are complementary to current protocols in screening for and diagnosis of COVID-19.

In addition to upper respiratory tract clinical samples, we profiled datasets derived from deep within the respiratory tract, comprising a dataset derived from COVID-19 patient lung biopsies as well as a separate dataset of bronchoalveolar lavage fluid (BALF) samples from COVID-19 patients; we analyzed these MS-data against our 639 peptide panel to determine whether our methodology was suitable for SARS-CoV-2 detection in these samples. We found a lack of high-confidence

peptides with high-quality spectra in these samples, with only a single MS run from the PXD022085 sample yielding the peptide FLALCADSIIGGAK which was not found in the datasets derived from higher up in the respiratory tract. Our results would suggest that samples collected using invasive methods (biopsy, lung fluid extraction), in addition to being taxing on the patients to collect, demonstrate insufficient concentrations of viral particles to be robustly detected using MS-based methods and the workflows presented here. The complexity of the sample matrices may also affect the ability to detect SARS-CoV-2 peptides, as the upper respiratory tract dataset which showed the fewest proposed target peptides- PXD023016- was also the only upper respiratory tract dataset which utilized viral transport medium in the collection of patient samples. Viral transport medium contains added serum as a part of its formulation, adding to the protein background of the collected samples. The deep lung datasets were also noted for their complexity, being either homogenized bulk lung tissue (PXD018094) or protein- and lipid-rich bronchoalveolar lavage fluid (PXD02085). In addition, the deep lung datasets had more sample preparation steps than the upper respiratory tract datasets, providing more opportunities for adding confounding variables to the analysis. Our results suggest that samples collected using minimally invasive methods from the upper respiratory tract (oropharyngeal/nasopharyngeal swabs and gargling samples) and using simplified, streamlined sample preparations would be most suitable for reliable detection of the SARS-CoV-2 virus targeting the high-confidence peptides we identify here—offering an optimal method for high-throughput diagnosis of infection.

While we believe the peptides presented here constitute promising targets for COVID-19 diagnosis, there are further experiments required to establish targeted proteomics as a viable methodology for detection of SARS-CoV-2 infection. The limits of detection of these peptides need to be reliably established in larger numbers of human samples collected in the clinic to determine the minimal number of viral particles that can be detected. This could help determine the optimal sample type and procedure for collection to ensure reliable results. In addition, proteomic analysis of samples collected at different stages of SARS-CoV-2 infection should be performed to determine viability of targeted proteomics for detection during the full life cycle of infection. Finally, the sample processing that accompanies bottom-up proteomics [58] should be optimized to be performed on a rapid time scale. Most conventional bottom-up proteomics experiments utilize trypsin digestions which occur overnight with incubation at

37 °C, meaning a single sample would have to be processed and analyzed over the course of two days; this would have to be significantly reduced as the conventional 24–48 h complete turnaround of RT-qPCR assays is being decreased through the use of strategies such as direct RT-qPCR [12], RT-LAMP [13], and CRISPR-based amplification strategies [59–61]. The turnaround time of clinical proteomics can potentially be decreased for individual samples using modified or alternative protein digestion enzymes with higher rates of reactivity [62]; in addition, automation of clinical proteomics technology can provide reproducible, robust analyses of patient samples [63, 64].

In addition to peptides derived empirically from clinical and in vitro datasets, we also included theoretical SARS-CoV-2 peptides predicted bioinformatically by Orsburn et al. [38] in our panel for validation; in doing so we were able to validate eight peptides in both clinical and in vitro datasets. It is worth noting, however, that of these eight peptides only two peptides were observed to have good quality spectra in the clinical data, supporting the need for caution in accepting peptide identifications. The validation workflow presented here was also able to identify peptides in mass spectrometry data which conventional unbiased algorithms, such as our database search workflow presented in Fig. 2b, are unable to identify; this may be of use in the analysis of complex patient and environmental mass spectrometry data collected for alternate purposes in the detection of SARS-CoV-2 under various conditions.

In conclusion, we interrogated multiple proteomic datasets from COVID-19 patients and in vitro experiments using bioinformatics workflows in order to determine which peptides from SARS-CoV-2 would make suitable targets for a clinical proteomics assay and which would make poor targets, potentially resulting in false negatives. Through our analyses, we found that of the 639 peptides that are readily detected across all samples, 87 of these were found to have a specific match to the SARS-CoV-2 proteome, rather than within the human proteome or other coronavirus proteomes. These peptides were narrowed down to 4 high-confidence peptides with excellent quality spectra found across most of the upper-respiratory tract clinical datasets analyzed in this study which we believe would be ideal candidates for diagnosis of COVID-19 via targeted proteomics. The workflows employed here for peptide identification and validation are well-documented, open-source, and hosted on the publicly accessible Galaxy Europe platform (usegalaxy.eu) where they can be edited, modified, or interfaced with other relevant bioinformatics tools to aid in analysis of proteomics data.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12014-021-09321-1>.

Additional file 1. Supplementary Data S1 with SearchGUI and PepQuery search parameters; supplementary figures 1–4, and supplementary tables 1–2.

Acknowledgements

We would like to thank the European Galaxy team and ELIXIR-Europe for the help in the support during Galaxy implementation and hosting the COVID-19 project webpage. We greatly appreciate inputs and help in data organization by Ms. Emma Leith.

Authors' contributions

ATR and SM drafted the original manuscript, performed data analyses and revised the final manuscript. DDAN contributed to data analysis. BAG, JEJ and TMG helped in software implementation and reviewing the manuscript. TJG critically reviewed the manuscript and provided supervision and funding for the project. PDJ conceptualized and supervised the project along with data analysis and manuscript preparation and review. All authors read and approved the final manuscript.

Authors' information

ATR and SM drafted the original manuscript, performed data analyses and revised the final manuscript. DDAN contributed to data analysis. BAG, JEJ and TMG helped in software implementation and reviewing the manuscript. TJG critically reviewed the manuscript and provided supervision and funding for the project. PDJ conceptualized and supervised the project along with data analysis and manuscript preparation and review. All authors read and approved the final manuscript.

Funding

We acknowledge funding for this work from the grant National Cancer Institute – Informatics Technology for Cancer Research (NCI-ITCR) grant 1U24CA199347 to T.J.G. 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.

Availability of data and materials

Data and workflow folder: <https://doi.org/10.5281/zenodo.4716149>

Data	Lab	Proteome exchange ID	Data availability
Cell culture	Gouveia et al	PXD018804	https://COVID19.galaxyproject.org/proteomics/PXD018804/
	Grenga et al	PXD018594	https://COVID19.galaxyproject.org/proteomics/PXD018594/
	Matthews et al	PXD018241	https://COVID19.galaxyproject.org/proteomics/PXD018241/

Data	Lab	Proteome exchange ID	Data availability
Clinical samples	Cardozo et al	PXD021328	https://COVID19.galaxyproject.org/proteomics/PXD019119/
	Ihling et al	PXD019423	https://COVID19.galaxyproject.org/proteomics/PXD018682/
	Rivera et al	PXD020394	https://COVID19.galaxyproject.org/proteomics/PXD020394/
	Leng et al	PXD018094	https://COVID19.galaxyproject.org/proteomics/PXD018094/
	Zeng et al	PXD022085	https://COVID19.galaxyproject.org/proteomics/PXD022085/
	Cardozo et al	PXD025214	https://COVID19.galaxyproject.org/proteomics/PXD025214/
	Bankar et al	PXD023016	https://COVID19.galaxyproject.org/proteomics/PXD023016/

Declarations

Ethics approval and consent to participate

As this study utilized open-source mass spectrometry data of anonymous human subjects from previous studies, no ethics approval or consent to participate was needed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biochemistry, Molecular and Cell Biology Building, University of Minnesota, 420 Washington Ave SE 7-129, Minneapolis, MN 55455, USA.

²Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455, USA. ³Department of Computer Science, University of Freiburg, Freiburg, Germany.

Received: 1 March 2021 Accepted: 1 May 2021

Published online: 10 May 2021

References

- 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 GF, Tan W. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727–33.
- Nuzzo J, Moss W, Kahn J, Rutgow L et al. Have states flattened the curve? <https://coronavirus.jhu.edu/>. Accessed 5 May 2021.

3. Sanyaolu A, Okorie C, Marinkovic A, Patidar R, Younis K, Desai P, Hosen Z, Padda I, Mangat J, Altaf M. Comorbidity and its impact on patients with COVID-19. *SN Comprehens Clin Med*. 2020;2(8):1069–76.
4. Ye Q, Wang B, Mao J. The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. *J Infect*. 2020;80(6):607–13.
5. Beigel JH, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, Hohmann E, Chu HY, Luetkemeyer A, Kline S, de Lopez Castilla D, Finberg RW, Dierberg K, Tapson V, Hsieh L, Patterson TF, Paredes R, Sweeney DA, Short WR, Touloumi G, Lye DC, Ohmagari N, Oh MD, Ruiz-Palacios GM, Benfield T, Fätkenheuer G, Kortepeter MG, Atmar RL, Creech CB, Lundgren J, Babiker AG, Pett S, Neaton JD, Burgess TH, Bonnett T, Green M, Makowski M, Osinusi A, Nayak S, Lane HC. Remdesivir for the treatment of Covid-19—final report. *New Engl J Med*. 2020;383(19):1813–26.
6. Riva L, Yuan S, Yin X, Martin-Sancho L, Matsunaga N, Pache L, Burgstaller-Muehlbacher S, De Jesus PD, Teriete P, Hull MV. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature*. 2020;586(7827):113–9.
7. Poland GA, Ovsyannikova IG, Crooke SN, Kennedy RB. SARS-CoV-2 vaccine development: current status. *Mayo Clin Proc*. 2020;95(10):2172–88.
8. Dagotto G, Yu J, Barouch DH. Approaches and challenges in SARS-CoV-2 vaccine development. *Cell Host Microbe*. 2020. <https://doi.org/10.1016/j.chom.2020.08.002>.
9. Jackson LA, Anderson EJ, Roupael NG, Roberts PC, Makhene M, Coler RN, McCullough MP, Chappell JD, Denison MR, Stevens LJ, Pruijssers AJ, McDermott A, Flach B, Doria-Rose NA, Corbett KS, Morabito KM, O'Dell S, Schmidt SD, Swanson PA, Padilla M, Mascola JR, Neuzil KM, Bennett H, Sun W, Peters E, Makowski M, Albert J, Cross K, Buchanan W, Pikaart-Tautges R, Ledgerwood JE, Graham BS, Beigel JH. An mRNA vaccine against SARS-CoV-2—preliminary report. *N Engl J Med*. 2020;383(20):1920–31.
10. Nagura-Ikeda M, Imai K, Tabata S, Miyoshi K, Murahara N, Mizuno T, Horiuchi M, Kato K, Imoto Y, Iwata M, Mimura S, Ito T, Tamura K, Kato Y. Clinical evaluation of self-collected saliva by quantitative reverse transcription-PCR (RT-qPCR), Direct RT-qPCR, reverse transcription-loop-mediated isothermal amplification, and a rapid antigen test to diagnose COVID-19. *J Clin Microbiol*. 2020;58(9):e01438-e1520.
11. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, Bleicker T, Brünink S, Schneider J, Schmidt ML, Mulders DG, Haagmans BL, van der Veer B, van den Brink S, Wijsman L, Goderski G, Romette J-L, Ellis J, Zambon M, Peiris M, Goossens H, Reusken C, Koopmans MP, Drosten C. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*. 2020;25(3):2000045.
12. Kriegova E, Fillerova R, Kvapil P. Direct-RT-qPCR detection of SARS-CoV-2 without RNA extraction as part of a COVID-19 testing strategy: from sample to result in one hour. *Diagnostics*. 2020;10(8):605.
13. Lu R, Wu X, Wan Z, Li Y, Jin X, Zhang C. A novel reverse transcription loop-mediated isothermal amplification method for rapid detection of SARS-CoV-2. *Int J Mol Sci*. 2020;21(8):2826.
14. Jung J, Ladha A, Saito M, Kim N-G, Woolley AE, Segel M, Barretto RPJ, Ranu A, Macrae RK, Faure G, Ioannidi EI, Krajcski RN, Bruneau R, Huang M-LW, Yu XG, Li JZ, Walker BD, Hung DT, Greninger AL, Jerome KR, Gootenberg JS, Abudayyeh OO, Zhang F. Detection of SARS-CoV-2 with SHERLOCK One-Pot testing. *N Engl J Med*. 2020;383(15):1492–4.
15. Tichopad A, Didier A, Pfaffl MW. Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants. *Mol Cell Probes*. 2004;18(1):45–50.
16. Schrader C, Schielke A, Ellerbroek L, Johne R. PCR inhibitors – occurrence, properties and removal. *J Appl Microbiol*. 2012;113(5):1014–26.
17. Foster MW, Gerhardt G, Robitaille L, Plante P-L, Boivin G, Corbeil J, Moseley MA. Targeted proteomics of human metapneumovirus in clinical samples and viral cultures. *Anal Chem*. 2015;87(20):10247–54.
18. Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc Natl Acad Sci*. 2007;104(14):5860–5.
19. Guerin M, Gonçalves A, Toiron Y, Baudelet E, Pophillat M, Granjeaud S, Fourquet P, Jacot W, Tarpin C, Sabatier R, Agavnian E, Finetti P, Adelaide J, Birnbaum D, Ginestier C, Charafe-Jauffret E, Viens P, Bertucci F, Borg J-P, Camoin L. Development of parallel reaction monitoring (PRM)-based quantitative proteomics applied to HER2-Positive breast cancer. *Oncotarget*. 2018;9(73):33762–77.
20. Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, Old WM, Cheung HT, Russell S, Wattawa JL, Goehle GR, Knight RD, Ahn NG. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal Chem*. 2004;76(13):3556–68.
21. Wu F-X, Gagné P, Droit A, Poirier GG. Quality assessment of peptide tandem mass spectra. *BMC Bioinform*. 2008;9(S6):S13.
22. Gouveia D, Grenga L, Gaillard JC, Gallais F, Bellanger L, Pible O, Armen-gaud J. Shortlisting SARS-CoV-2 peptides for targeted studies from experimental data-dependent acquisition tandem mass spectrometry data. *Proteomics*. 2020;20(14):e2000107.
23. Grenga L, Gallais F, Pible O, et al. Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines. *Emerging Microbes Infect*. 2020;9(1):1712–21.
24. Davidson AD, Williamson MK, Lewis S, et al. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. *Genome Biol*. 2020;21(6):68.
25. Cardozo KHM, Lebkuchen A, Okai GG, Schuch RA, Viana LG, Olive AN, Lazari CDS, Fraga AM, Granato CFH, Pintão MCT, Carvalho VM. Establishing a mass spectrometry-based system for rapid detection of SARS-CoV-2 in large clinical sample cohorts. *Nat Commun*. 2020;11(1):6201–6201.
26. Ihling C, Tänzler D, Hagemann S, Kehlen A, Hüttelmaier S, Arlt C, Sinz A. Mass spectrometric identification of SARS-CoV-2 proteins from gargle solution samples of COVID-19 patients. *J Proteome Res*. 2020;19(11):4389–92.
27. Rivera B, Leyva A, Portela MM, Moratorio G, Moreno P, Durán R, Lima A. Quantitative proteomic dataset from oro- and naso-pharyngeal swabs used for COVID-19 diagnosis: Detection of viral proteins and host's biological processes altered by the infection. *Data Brief*. 2020;32:106121.
28. Zeng HL, Chen D, Yan J, Yang Q, Han QQ, Li SS, Cheng L. Proteomic characteristics of bronchoalveolar lavage fluid in critical COVID-19 patients. *FEBS J*. 2020. <https://doi.org/10.1111/febs.15609>.
29. Leng L, Cao R, Ma J, Mou D, Zhu Y, Li W, Lv L, Gao D, Zhang S, Gong F, Zhao L, Qiu B, Xiang H, Hu Z, Feng Y, Dai Y, Zhao J, Wu Z, Li H, Zhong W. Pathological features of COVID-19-associated lung injury: a preliminary proteomics report based on clinical samples. *Signal Transduct Targeted Therapy*. 2020;5(1):1–9.
30. Bankar R, Suvarna K, Ghantasala S, Banerjee A, Biswas D, Choudhury M, Palanivel V, Salkar A, Verma A, Singh A, Mukherjee A, Pai MGJ, Roy J, Srivastava A, Badaya A, Agrawal S, Shrivastav O, Shastri J, Srivastava S. Proteomic investigation reveals dominant alterations of neutrophil degranulation and mRNA translation pathways in patients with COVID-19. *iScience*. 2021;24(3):102135.
31. Wen B, Wang X, Zhang B. PepQuery enables fast, accurate, and convenient proteomic validation of novel genomic alterations. *Genome Res*. 2019;29(3):485–93.
32. McGowan T, Johnson JE, Kumar P, Sajulga R, Mehta S, Jagtap PD, Griffin TJ. Multi-omics visualization platform: an extensible galaxy plug-in for multi-omics data visualization and exploration. *GigaScience*. 2020;9(4):gia025.
33. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinform*. 2009;10:1–9.
34. Saunders JK, Gaylord DA, Held NA, Symmonds N, Dupont CL, Shepherd A, Kinkade DB, Saito MA. METATRYP v 20: Metaproteomic least common ancestor analysis for taxonomic inference using specialized sequence assemblies—standalone software and web servers for marine microorganisms and coronaviruses. *J Proteome Res*. 2020;19(11):4718–29.
35. Cardozo KHM, Lebkuchen A, Okai GG, et al. Establishing a mass spectrometry-based system for rapid detection of SARS-CoV-2 in large clinical sample cohorts. *Nat Commun*. 2020;11(1):6201.
36. Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: an open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics*. 2011;11(5):996–9.
37. Vaudel M, Burkhardt JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, Martens L, Barsnes H. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat Biotechnol*. 2015;33(1):22–4.
38. Orsburn BC, Jenkins C, Miller SM, Neely BA, Bumpus NN. In silico approach toward the identification of unique peptides from viral protein infection: Application to COVID-19. *Cold Spring Harbor Laboratory*; 2020. CR-MEDICINE-D-20-00111.
39. Li K, Vaudel M, Zhang B, Ren Y, Wen B. PDV: an integrative proteomics data viewer. *Bioinformatics*. 2019;35(7):1249–51.

40. Gurdeep Singh R, Tanca A, Palomba A, Van Der Jeugt F, Verschaffelt P, Uzzau S, Martens L, Dawyndt P, Mesuere B. Unipept 4.0: functional analysis of metaproteome data. *J Proteome Res*. 2019;18(2):606–15.
41. Capobianchi MR, Rueca M, Messina F, Giombini E, Carletti F, Colavita F, Castilietti C, Lalle E, Bordi L, Vairo F, Nicastrì E, Ippolito G, Gruber CEM, Bartolini B. Molecular characterization of SARS-CoV-2 from the first case of COVID-19 in Italy. *Clin Microbiol Infect*. 2020;26(7):954–6.
42. Colavita F, Lapa D, Carletti F, Lalle E, Messina F, Rueca M, Matusali G, Meschi S, Bordi L, Marsella P. In Virological characterization of the first 2 COVID-19 patients diagnosed in Italy: phylogenetic analysis, virus shedding profile from different body sites, and antibody response kinetics. *Open Forum Infectious Diseases*, Oxford: Oxford University Press US; 2020. p. ofaa403.
43. Pozniak Y, Balint-Lahat N, Rudolph JD, Lindskog C, Katzir R, Avivi C, Pontén F, Ruppén E, Barshack I, Geiger T. System-wide clinical proteomics of breast cancer reveals global remodeling of tissue homeostasis. *Cell Syst*. 2016;2(3):172–84.
44. Yanovich G, Agmon H, Harel M, Sonnenblick A, Peretz T, Geiger T. Clinical proteomics of breast cancer reveals a novel layer of breast cancer classification. *Can Res*. 2018;78(20):6001–10.
45. Chen Y, Huang A, Ao W, Wang Z, Yuan J, Song Q, Wei D, Ye H. Proteomic analysis of serum proteins from HIV/AIDS patients with *Talaromyces marneffei* infection by TMT labeling-based quantitative proteomics. *Clin Proteomics*. 2018;15(1):40.
46. Messner CB, Demichev V, Wendisch D, Michalick L, White M, Freiwald A, Textoris-Taube K, Vernardis SI, Egger A-S, Kreidl M. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Syst*. 2020;11(1):11–24. e4.
47. Wikramaratna PS, Paton RS, Ghafari M, Lourenço J. Estimating the false-negative test probability of SARS-CoV-2 by RT-PCR. *Euro Surveill*. 2020;25(50):2000568.
48. Shteynberg D, Nesvizhskii AI, Moritz RL, Deutsch EW. Combining results of multiple search engines in proteomics. *Mol Cell Proteom MCP*. 2013;12(9):2383–93.
49. Liu W-K, Xu D, Xu Y, Qiu S-Y, Zhang L, Wu H-K, Zhou R. Protein profile of well-differentiated versus un-differentiated human bronchial/tracheal epithelial cells. *Heliyon*. 2020;6(6):e04243.
50. Marcus-Sekura C. Process changes and their effect on process evaluation for viral clearance. *Dev Biol Stand*. 1996;88:125–30.
51. Chang C-K, Sue S-C, Yu T-H, Hsieh C-M, Tsai C-K, Chiang Y-C, Lee S-J, Hsiao H-H, Wu W-J, Chang W-L. Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci*. 2006;13(1):59–72.
52. Chang C-K, Hou M-H, Chang C-F, Hsiao C-D, Huang T-H. The SARS coronavirus nucleocapsid protein—forms and functions. *Antiviral Res*. 2014;103:39–50.
53. de Haan CA, Rottier PJ. Molecular interactions in the assembly of coronaviruses. *Adv Virus Res*. 2005;64:165–230.
54. Mortola E, Roy P. Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system. *FEBS Lett*. 2004;576(1–2):174–8.
55. Bar-On YM, Flamholz A, Phillips R, Milo R. Science Forum: SARS-CoV-2 (COVID-19) by the numbers. *Elife*. 2020;9:e57309.
56. Hiscox JA, Wurm T, Wilson L, Britton P, Cavanagh D, Brooks G. The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus. *J Virol*. 2001;75(1):506–12.
57. Ravi N, Cortade DL, Ng E, Wang SX. Diagnostics for SARS-CoV-2 detection: A comprehensive review of the FDA-EUA COVID-19 testing landscape. *Biosensors Bioelectron*. 2020;165:112454.
58. Gundry RL, White MY, Murray CI, Kane LA, Fu Q, Stanley BA, Van Eyk JE. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr Protocols Mol Biol*. 2010;90(1):10–25.
59. Joung J, Latha A, Saito M, Segel M, Bruneau R, Huang M-LW, Kim N-G, Yu X, Li J, Walker BD, Greninger AL, Jerome KR, Gootenberg JS, Abudayyeh OO, Zhang F. Point-of-care testing for COVID-19 using SHERLOCK diagnostics. *medRxiv*. 2020. <https://doi.org/10.1101/2020.05.04.20091231>.
60. Dara M, Talebzadeh M. CRISPR/Cas as a Potential Diagnosis Technique for COVID-19. *Avicenna J Med Biotechnol*. 2020;12(3):201–2.
61. Zhang F, Abudayyeh OO, Gootenberg JS. A protocol for detection of COVID-19 using CRISPR diagnostics. *A protocol for detection of COVID-19 using CRISPR diagnostics*. 2020, 8.
62. Gutierrez DB, Gant-Branum RL, Romer CE, Farrow MA, Allen JL, Dahal N, Nei Y-W, Codreanu SG, Jordan AT, Palmer LD. An integrated, high-throughput strategy for multiomic systems level analysis. *J Proteome Res*. 2018;17(10):3396–408.
63. Lee J, Kim H, Sohn A, Yeo I, Kim Y. Cost-effective automated preparation of serum samples for reproducible quantitative clinical proteomics. *J Proteome Res*. 2019;18(5):2337–45.
64. Müller T, Kalxdorf M, Longuespée R, Kazdal DN, Stenzinger A, Krijgsveld J. Automated sample preparation with SP 3 for low-input clinical proteomics. *Mol Syst Biol*. 2020;16(1):9111.
65. Rasche H, Hiltmann S. Galactic Circos: User-friendly Circos plots within the Galaxy platform. *Giga Science*. 2020;9(6):giaa065.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

