

LC-MS for detection of SARS-CoV-2 viral and host proteins

Authors: Kruthi Suvarna¹,
Medha Gayathri J Pai¹, Sanjeeva Srivastava¹,
Debadeep Bhattacharyya², Kerry Hassell²

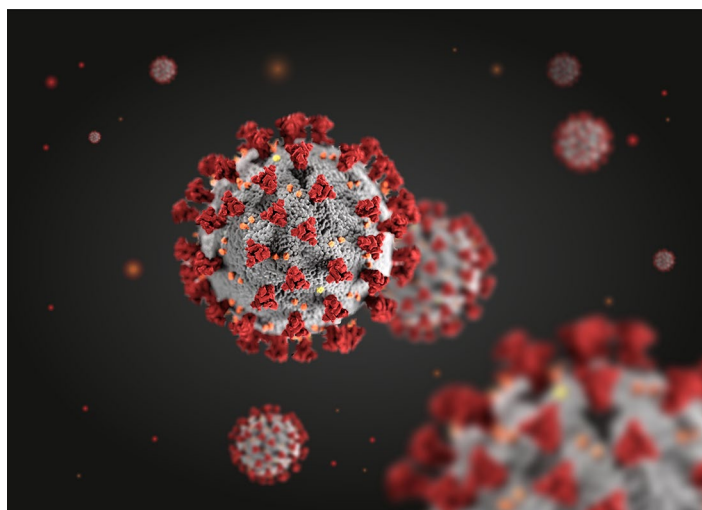
¹Indian Institute of Technology, Bombay,
Powai, Mumbai, India

²Thermo Fisher Scientific,
San Jose, CA, USA

Keywords: COVID-19, SARS-CoV-2,
host response, nasopharyngeal, plasma,
mass spectrometry, omics, TSQ Altis,
Orbitrap Fusion Lumos, Vanquish Flex UHPLC
system, EASY-nLC 1200, molecular pathways,
prognostic biomarkers, proteomics

Goal

To develop a comprehensive LC-MS assay for monitoring and quantifying targeted peptide surrogates from critical proteins from nasopharyngeal swab (NPS) and plasma samples obtained from individuals testing positive for SARS-CoV-2 coronavirus with reverse transcription-polymerase chain reaction (RT-PCR) based analysis. Identification of viral and host peptide surrogates from critical proteins was conducted via simple extraction processes followed by analysis using high-resolution mass spectrometry. The surrogate peptide list was transitioned to a targeted SRM method performed on a triple quadrupole mass spectrometer to determine if the relative expression ratios between hosts were defined by expected viral load. It is the intent of this study to find if the differentially expressed proteins identified correlate with the viral load



of the host, which would indicate if these proteins can be effective candidates for potential prognostic biomarkers of COVID-19 severity prediction. This study will try and reveal if mass spectrometry-based peptide tests have potential to be efficiently used by clinicians for diagnosis and host marker monitoring for disease progression.

Introduction

Detection of exposure to viral infection is often conducted by either determining viral load by real-time RT-PCR or monitoring virus-resisting antibodies generated within the human body via antigen tests. For COVID-19, RT-PCR has been the gold standard for detection of viral load in the human body. To date, millions of RT-PCR tests have been carried out for detection of COVID-19 in almost every country. As commonly observed for RT-PCR and

such similar, fast tests, cross-reactions are very common. Considering the severity of the disease, there is always a need for a complimentary technology that can confirm the outcome of the results, especially results considered on the threshold. Such results can often turn out to be false positive or false negative.

Nasopharyngeal swab sampling offers a very high viral titer (also known as viral load) and provides a high concentration of virus in a specified volume of liquid. It is one of the easiest, most economical, and non-invasive ways of sample acquisition and is currently being used for RT-PCR-based detection of the SARS-CoV-2 virus. A deeper understanding of the viral proteins that generate a high impact to the human proteomes is required to respond in a particular way to such viral infections. This not only helps in understanding the extent of the infection, but also reduces the number of false positives and false negatives. To address such key issues, proteomics-based investigations for the identification of biomarkers for COVID-19 have already begun among the global scientific community.

In this study, a comprehensive targeted protein quantification-based investigation was performed to identify surrogate peptides from SARS-CoV-2 viral and host proteins from respiratory and plasma specimens of COVID-19 positive individuals using high resolution accurate mass (HRAM) mass spectrometry. Depending on the clinical symptoms, positive patients as advised by clinicians were further grouped into severe (patients requiring mechanical ventilation and having severe symptoms of acute respiratory distress or bilateral pneumonia) and non-severe (patients having mild symptoms of cough, fever, fatigue, and breathlessness without invasive ventilation) groups. We identified several differentially expressed proteins from host and pathogen that correlated with the severity of the symptoms using label-free mass spectrometry-based proteomics, as well as targeted approaches, such as selective reaction monitoring (SRM). The host proteome analysis identified neutrophil degranulation, interleukin pathway, and mRNA translation of proteins as major pathways that were modulated in the COVID-19-infected individuals, thus providing the landscape of COVID-19 pathophysiology. This report provides a comprehensive application of LC-MS for the detection and monitoring of COVID-19 infection severity amongst the population for clinical research.

Experimental

Plasma proteome analysis

Sample preparation

Approximately 1 mL of human blood was collected and centrifuged at 3,000 rpm for 10 minutes to separate the plasma. The separated plasma was then incubated at 56 °C for 30 min for viral inactivation and further stored at -80 °C in cryovials. The plasma was depleted, reduced with Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) to a final concentration of 20 mM at 37 °C for 1 hour, and then alkylated with iodoacetamide (final concentration 40 mM) for 15 min protected from light. Proteins were subjected to enzymatic digestion by trypsin at an enzyme/substrate ratio of 1:30 for 16 hours at 37 °C.

Liquid chromatography

For initial detection assays, 1 µg peptide of each sample was separated with a Thermo Scientific™ EASY-nLC™ 1200 system with a gradient of 80% acetonitrile (ACN) and 0.1% formic acid (FA) for 120 min with blank injections after every sample. Bovine serum albumin (BSA) was run at the starting and endpoint of each set of the run to ensure instrument performance was stable and reliable. All samples were loaded onto the LC column at a flow rate of 300 nL/min.

Targeted analyses of identified peptides were performed on the Thermo Scientific™ Vanquish™ Flex Quaternary UHPLC system. The peptides were separated using a Thermo Scientific™ Hypersil GOLD™ C18 column (1.9 µm, 100 × 2.1 mm) using a flow rate of 450 µL/min for a total time of 10 minutes. The eluents used were 0.1% FA in water as the mobile phase A and 80% ACN in 0.1% FA as the mobile phase B.

Mass spectrometry

Initial proteomic discovery data was acquired on the Thermo Scientific™ Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer. Mass spectrometric data acquisition was done in data-dependent acquisition mode with a mass scan range of m/z 375–1,700 and a mass resolution of 60,000. A mass window of 10 ppm was set with a dynamic exclusion duration of 40 s. All MS/MS data was acquired with higher energy collision dissociation (HCD) fragmentation. Targeted SRM analyses were performed on the Thermo Scientific™ TSQ Altis™ triple quadrupole mass spectrometer. The method parameters for Orbitrap Fusion Lumos MS and TSQ Altis systems are listed in Tables 1 and 2, respectively.

Table 1. Orbitrap Fusion Lumos mass spectrometer settings for the analysis of COVID-19-infected plasma and nasopharyngeal swab samples

Parameter	Value
Source voltage (V)	2,000
Full Scan resolution	60,000
Scan range (m/z)	375–1700
Data-dependent MS/MS	MIPS
Include charge state	2–6
Dynamic exclusion (s)	40
Extraction mass window (ppm)	10

Table 2. TSQ Altis mass spectrometer settings for the analysis of COVID-19-infected plasma and nasopharyngeal swab samples

Parameter	Value			
Ion source type	NSI			
Spray voltage	Static			
Positive ion (V)	2,100			
Negative ion (V)	600			
Sweep gas (Arb)	0			
Ion transfer tube temp. (°C)	300			
LC gradient for UHPLC coupled with TSQ Altis MS				
Gradient	<i>Time</i>	<i>Flow</i>	<i>%B</i>	<i>Curve</i>
	<i>[mm:ss]</i>	<i>[μL/min]</i>		
	00:00	450	2	5
	06:00	450	45	5
	06:50	450	95	5
	07:00	450	95	5
	07:50	450	2	5
	10:00	450	2	5
The column compartment was kept at 40 °C.				
TSQ Altis MS settings				
Use cycle time	TRUE			
Cycle time (s)	2			
Use calibrated RF lens	TRUE			
Q1 resolution (FWHM)	0.7			
Q3 resolution (FWHM)	0.7			
CID gas argon (mTorr)	2.5			
Source fragmentation (V)	0			
Chromatographic peak width (s)	20			

Nasopharyngeal swab proteome analysis

Sample preparation

Nasopharyngeal swab samples were obtained from symptomatic adults. A sterile cotton swab was used to collect a clinical specimen and store it in a tube containing viral transport media (VTM) at 4 °C. Approximately 800 μL of the sample from the stored tube was dispensed into a sterile tube. The tube was incubated at 65 °C for 45 min for heat inactivation of the virus. 200 μL of inactivated swab sample collected in VTM were added

into a microcentrifuge tube containing 600 μL of acetone, ethanol, and isopropanol. All the tubes were kept at -20 °C for 4 h. Next, the microcentrifuge tubes were centrifuged at 15,000 g for 20 min at 4 °C. Any proteins present in the samples were precipitated into a pellet, while supernatants were discarded. Dried pellets in the microcentrifuge tubes were stored at -80 °C until they could be further processed. Approximately 75 μL of urea lysis buffer (8 M urea, 50 mM Tris pH 8.0, 75 mM NaCl, 1mM MgCl₂) were added to each protein pellet. From each sample, ~30 μg of protein was taken in a 1.5 mL tube for enzymatic digestion. For reducing the disulfide bonds in the proteins, 20 mM of tris (2-carboxyethyl) phosphine (TCEP) was added, and the tubes were incubated at 37 °C for 60 min. After reduction, iodoacetamide (IAA) was added to a final concentration of 37.5 mM alkylate and incubated for 20 min at room temperature, protected from light. The sample was diluted with a buffer containing 25 mM Tris pH 8.0 and 1 mM CaCl₂. Thermo Scientific™ Pierce™ trypsin protease was added at an enzyme to substrate ratio of 1:30, and tubes were incubated at 37 °C and shaken for 16 h. RT-PCR results were provided by the clinicians from the Kasturba hospital, Mumbai. See reference publications for further details.^{2,3}

Liquid chromatography

A desalting procedure is carried out to remove the salts from the urea buffer used for the lysis of the sample. In this procedure, the digested peptides are bound to in-house made C18 ziptips. After washing to remove the salts, the clean peptides are then eluted from the ziptips using acetonitrile. Desalted peptides were reconstituted in 0.1% (v/v) formic acid (FA). For separation, 0.1% FA was used to equilibrate the pre-column, Thermo Scientific™ Acclaim™ 100 C18 HPLC column (0.1 × 20 mm, 5 μm, P/N 164564), and analytical column, Thermo Scientific™ EASY-Spray™ C18 column (75 cm × 75 μm, P/N ES803A). Peptides (1 μg) were loaded onto the nanoLC column followed by separation along an LC gradient comprising two buffers, 80% ACN and 0.1% FA for 120 min at a flow rate of 300 nL/min using an EASY nLC 1200 system. The targeted analysis was performed on the Vanquish Flex UHPLC system using the Hypersil GOLD C18 column (100 × 2.1 mm, 1.9 μm, P/N 25002-102130-V) for the separation of the peptides. The separation temperature was 40 °C. The flow rate was 450 μL/min for a total run time of 10 min. The mobile phase was 0.1% FA in water for mobile phase A and 80% ACN with 0.1% FA for mobile phase B. The gradient used for chromatographic separation of the peptides is shown in Table 2.

Mass spectrometry

Nasopharyngeal swab mass spectrometry experimental details were identical to plasma parameters.

Data analysis

Raw datasets from the Orbitrap Fusion Lumos mass spectrometer were processed with MaxQuant software (Max-Planck-Institute of Biochemistry) against the Human Swiss-Prot database (Bairoch and Apweiler), searched with the built-in Andromeda Search Engine of MaxQuant. The proteins found to be of statistical significance and showing upregulation in COVID-19 severe versus non-severe and negative samples in the discovery data were selected and used for a targeted study. The list of transitions, covering 35 peptides from 13 proteins, was prepared for unique peptides of these selected proteins with Skyline software (Ver 20.2.1.286). This list included a heavy labeled spiked-in synthetic peptide essential for monitoring the consistency of the mass spectrometry runs.

Results

Plasma proteome analysis

Label-free quantification (LFQ) of 74 depleted plasma samples from COVID-19-infected individuals was performed. Figure 1 depicts the schematic workflow of label-free quantification under discovery proteomics. This study has also investigated the proteomic alteration between the non-severe and severe cohort, which provides a list of 25 significantly differentially expressed proteins. The proteins, which include kallistatin (SERPINA4), serum amyloid P-component (APCS), protein S100-A8 (S100A8), and fibrinogen gamma chain (FGG), were found to be up-regulated in the severe cohort, whereas proteins such as complement factor D (CFD), monocyte differentiation antigen (CD14), and complement component C8 alpha chain (C8A), were found to be down-regulated in the severe when compared to non-severe individuals. The violin plots for a few differentially expressed proteins (DEPs) are shown in Figure 2.

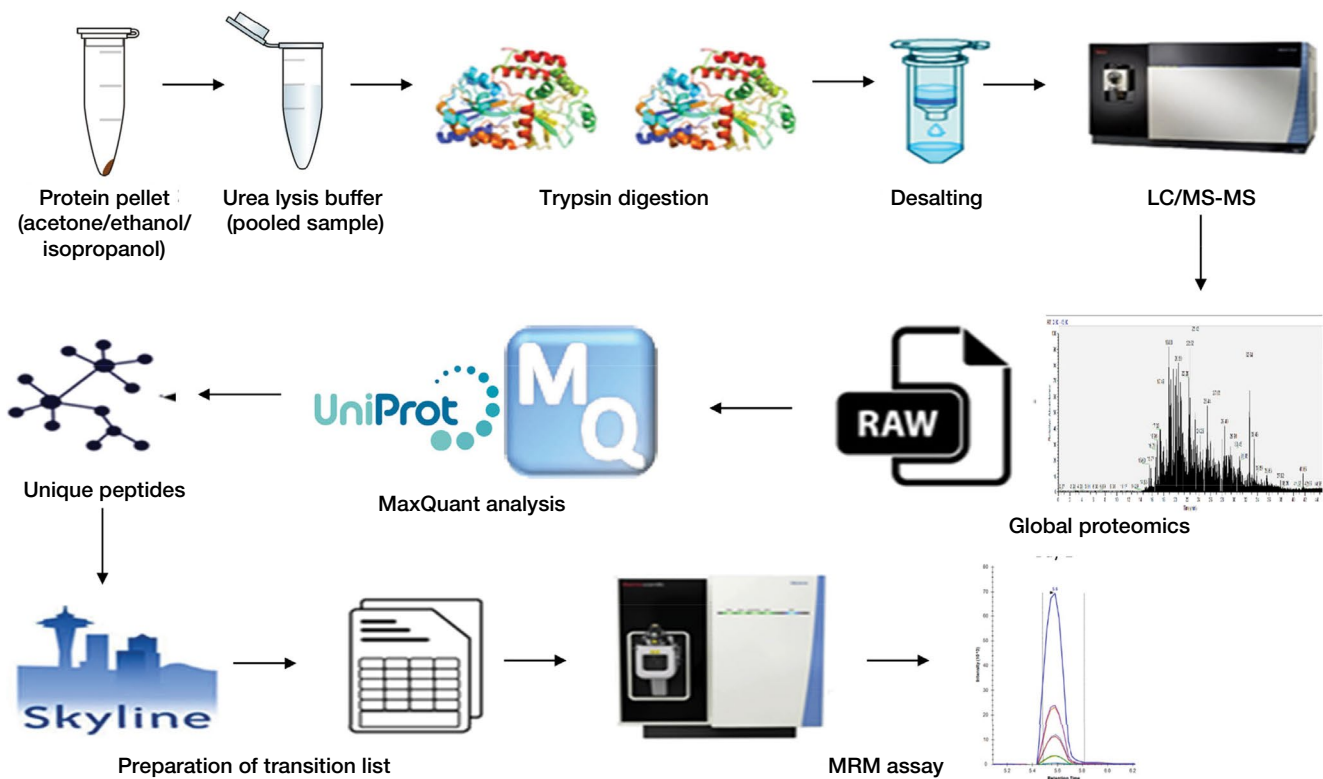


Figure 1. Workflow of mass spectrometry-based analysis of COVID-19 samples. For MS analysis, proteins in swab samples were precipitated using three organic solvents, acetone, ethanol, and isopropanol. For plasma samples, the depletion of high abundant plasma proteins was performed, and the samples were lysed using 6 M urea lysis buffer. The proteins were digested, desalted, and subjected to mass spectrometry analysis. Raw files from mass spectrometry were analyzed using MaxQuant software, and significant proteins were selected for MRM analysis.

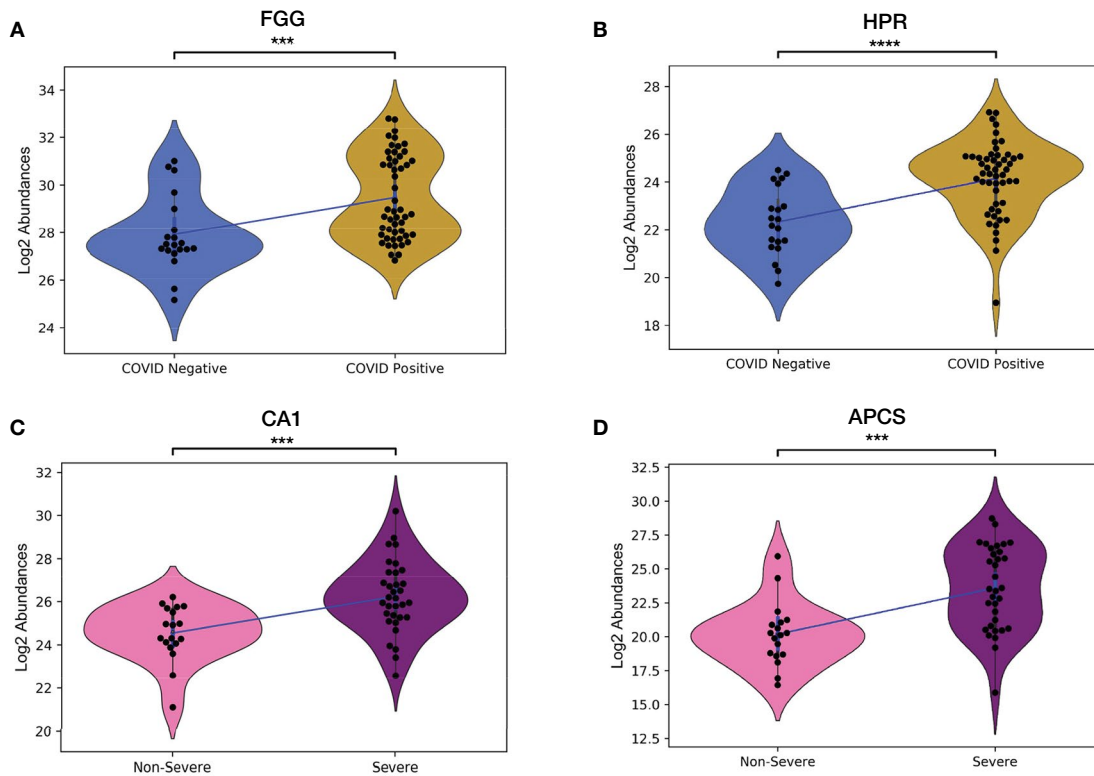


Figure 2. Significant differentially expressed proteins (DEPs) detected in the COVID-19 plasma samples. Violin plots for proteins Fibrinogen gamma chain (FGG) (A) and Haptoglobin-related protein (HPR) (B) in COVID-19 positive and COVID-19 negative Violin plots for proteins Carbonic anhydrase 1 (CA1) (C) and Serum amyloid P-component (APCS) (D) in COVID-19 severe and non-severe patient samples. (p value annotation: ***: $1.00e-04 < p \leq 1.00e-03$; ****: $p \leq 1.00e-04$). The “p” refers to the level of significance for the statistics used. The violin plots were made with Log2-transformed data, where the significance level was calculated based on Welch’s t-test independent samples with Bonferroni correction (p value annotation legend: ***: $1.00e04 < p \leq 1.00e-03$; ****: $p \leq 1.00e-04$).

Targeted SRM analysis

The SRM study aimed to validate the differentially regulated proteins found between samples obtained from severe and non-severe cohorts from the LFQ data. Synthetically heavy labeled peptides FEDGVLPDYPR and THZLYTHVZDAIK were spiked into all samples to ensure data quality. Figure 3 shows the instrument response for the spiked synthetic peptide. Based on the response of the differentially regulated peptides, the list of peptides was further refined to focus on peptides that exhibited significant dysregulation (adjusted p values below 0.05) between severe and non-severe cohorts. The signals were annotated, and transitions were refined matching with the

library to give DOTP values for all peptides. DOTP value is a measure of the match between the experimental peak of the peptide and its library fragmentation patterns. Thus, the refined list had 183 transitions belonging to 28 peptides of 9 host proteins and 1 synthetic peptide. The proteins AGT, APOB, SERPINA3, FGG, and SEPRING1 exhibited 3 or more peptides with a peak area change of more than 3 and adjusted p-value less than 0.05 at a confidence of 95–99%. Targeted SRM-based validation for a few host peptides is shown in Figure 4. This validates that, for the given set of samples, these proteins show statistically significant overexpression in COVID-19 severe individuals compared to COVID-19 non-severe individuals.

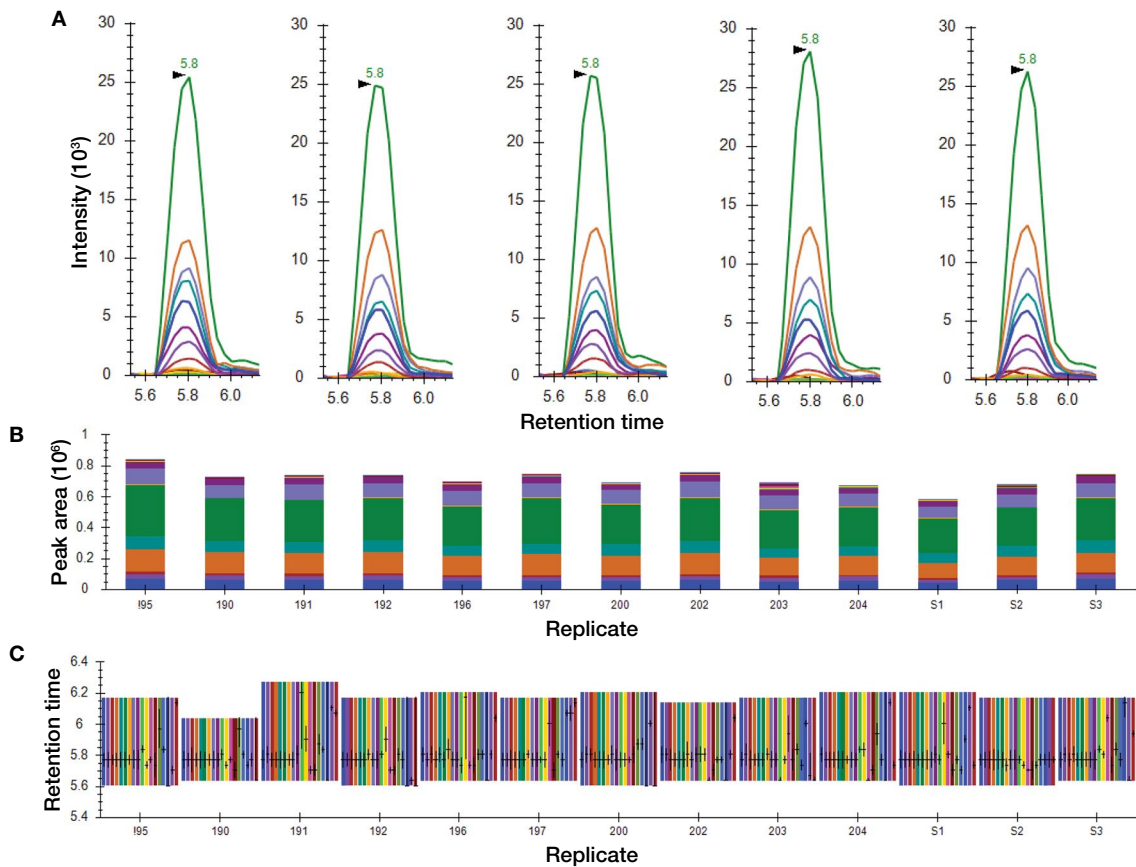


Figure 3. Response of heavy labeled synthetic peptide (FEDGCLDPDYPR) using a TSQ Altis mass spectrometer. Equal amounts of synthetic peptides are spiked into the samples. The peptide response was uniform in the different samples. This allows the monitoring of run-to-run variability and acts as an internal QC standard. Intensity (A), peak area (B), and retention time (C) for the spiked-in peptides.

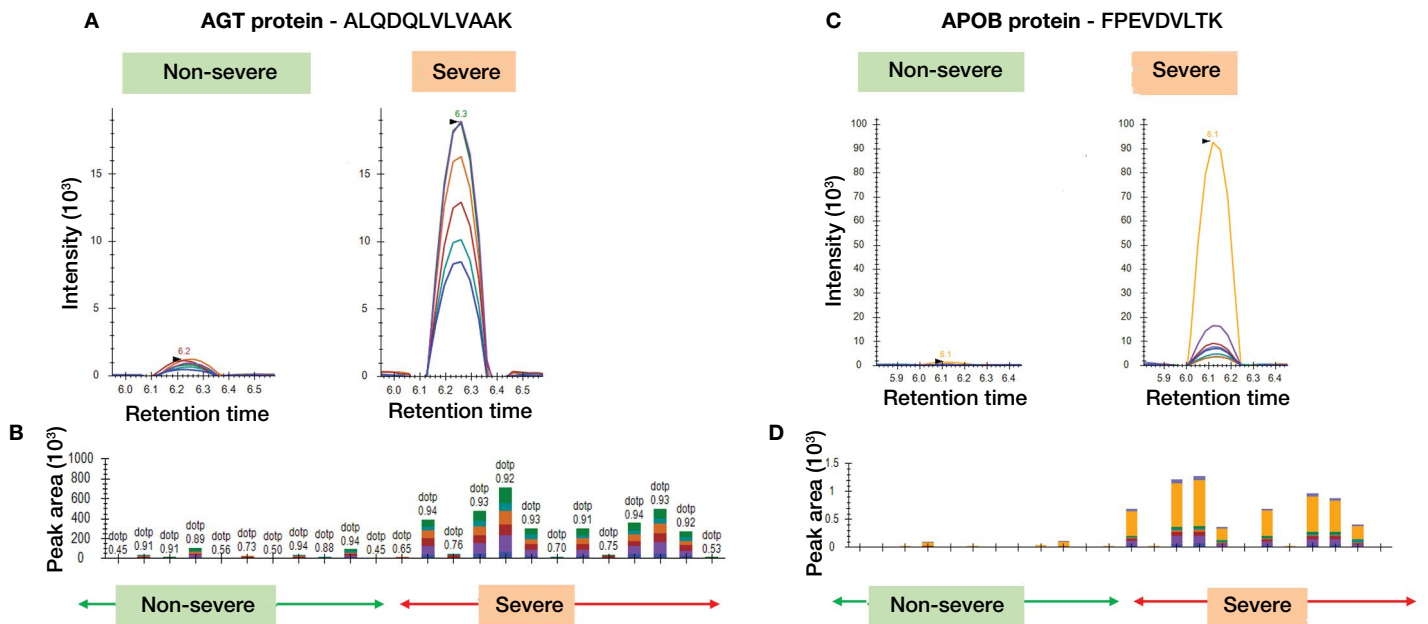


Figure 4. Targeted MRM-based validation of host peptides detected in the plasma samples from COVID-19-infected individuals. Intensity (A) and peak area (B) for the peptide specific for AGT protein detected in COVID-19 severe and non-severe patients. Intensity (C) and peak area (D) for the peptide specific for APOB protein detected in COVID-19 severe and non-severe individuals.

Nasopharyngeal swab analysis

Initially, shotgun mass spectrometry analysis was performed for swab samples in three different solvent conditions. However, a sample pool made from all the three solvents yielded maximum coverage of the peptides and was used for experiments. The raw dataset was processed using MaxQuant software against the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Swiss-Prot database. To further validate the viral peptides detected from the discovery set of 23 individuals, we selected 11 peptides and performed SRM assays on a cohort comprising 41 individuals. The SRM-based assay aimed at quantifying distinct viral peptides (NGSIHLYFDK, ELLQNGMNGR, DGHVETFYPK, DFMSLSEQLR) (P0DTD1) belonging to Replicase polyprotein 1ab; peptides (IQDLSSTASALGK, NTQEVEFAQVK, FLPFQQFGR, and QIAPGQTGK) belonging to SARS-CoV-2 spike glycoprotein (P0DTC2); and peptides (NSTPGSSR, GGSQASSR, DGIWVATEGALNTPK) belonging to SARS-CoV-2 nucleoprotein (P0DTC9) (Figures 5-7). Table 3 shows the list of viral peptides monitored.

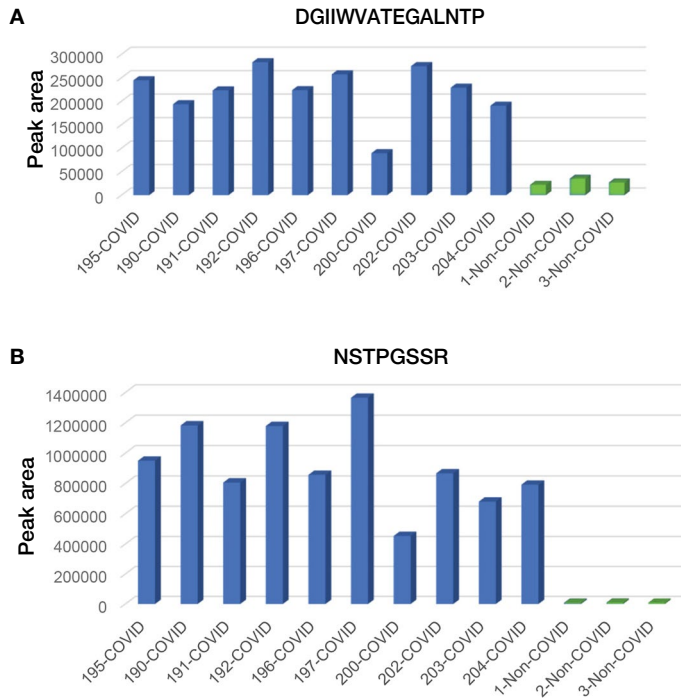


Figure 5. Peak area of peptides from SARS-CoV-2 nucleoprotein detected in swab samples of COVID-19-infected individuals and healthy individuals (non-COVID). Peak area of the peptides DGIWVATEGALNTPK (A) and NSTPGSSR (B) detected in swab samples of COVID-19-infected individuals and healthy individuals (non-COVID). The peak areas were obtained from the MRM data using Skyline software.

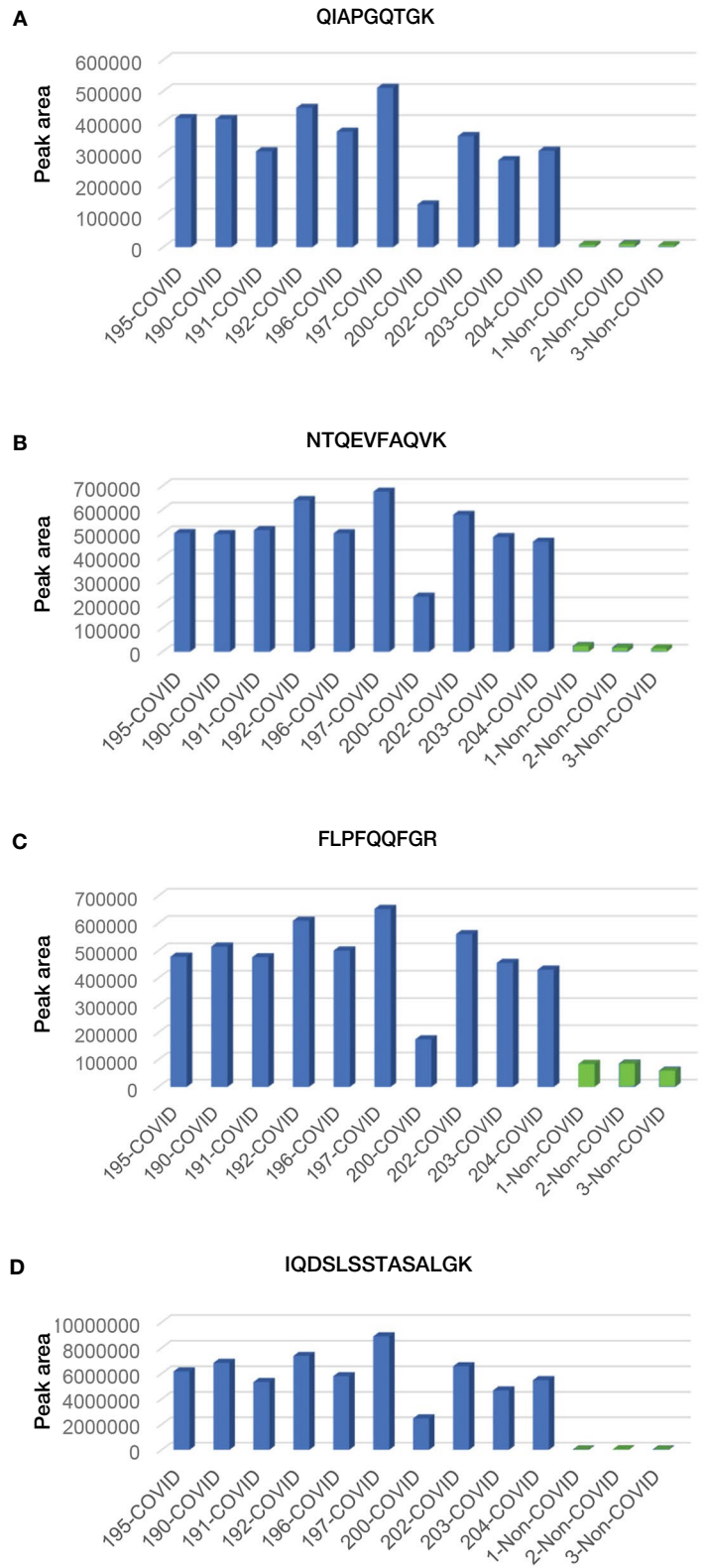


Figure 6. Peak area of peptides from SARS-CoV-2 spike glycoprotein detected in swab samples of COVID-19-infected individuals and healthy individuals (non-COVID). The peak areas were obtained from the MRM data using Skyline software.

Table 3. Viral peptides from different SARS-CoV-2 proteins detected in COVID +/- swab samples using MRM-based targeted proteomics assay using Skyline software.

Peptide	Protein accession	Protein name
DFMSLSEQLR	P0DTD1	sp P0DTD1 R1AB_SARS2
DGHVETFYFK	P0DTD1	sp P0DTD1 R1AB_SARS2
ELLQNGMNGR	P0DTD1	sp P0DTD1 R1AB_SARS2
NGSIHLYFDK	P0DTD1	sp P0DTD1 R1AB_SARS2
QIAPGQTGK	P0DTC2	sp P0DTC2 SPIKE_SARS2
FLPFQQFGR	P0DTC2	sp P0DTC2 SPIKE_SARS2
NTQEVFAQVK	P0DTC2	sp P0DTC2 SPIKE_SARS2
IQDSLSTASALGK	P0DTC2	sp P0DTC2 SPIKE_SARS2
DGIIVVATEGALNTPK	P0DTC9	sp P0DTC9 NCAP_SARS2
GGSQASSR	P0DTC9	sp P0DTC9 NCAP_SARS2
NSTPGSSR	P0DTC9	sp P0DTC9 NCAP_SARS2
FEDGVLDPDYPR		Synthetic peptide

We also performed a comprehensive proteomics-based investigation of nasopharyngeal swab samples from COVID-19-infected individuals to identify host prognosis markers. Figure 8 shows representative mass spectrometry generated chromatograms of swab samples. Few of the host proteins, such as interleukin-6, L-lactate dehydrogenase (LDH), C-reactive protein, ferritin, and aspartate aminotransferase, were found to be upregulated in COVID-19-infected individuals using targeted SRM study. The targeted SRM based validation for the peptide-specific to LDHA detected in the COVID-19-infected swab samples is shown in Figure 9A.

Standard curves of the heavy synthetic peptide HSGFEDELSEVLENQSSQAELK and light synthetic peptide DGIIVVATEGALNTPK were generated by preparing serial dilutions of the crude synthetic peptides. The crude light synthetic peptide was diluted in the range of 25 to 125 ng concentration. The standard curve for this peptide was plotted using the intensity of the peak area against the concentration of the peptide. The intensity of the peak area was proportional to the amount of the synthetic peptide. The lowest range of detection for this peptide was at 29 ng (Figure 9B).

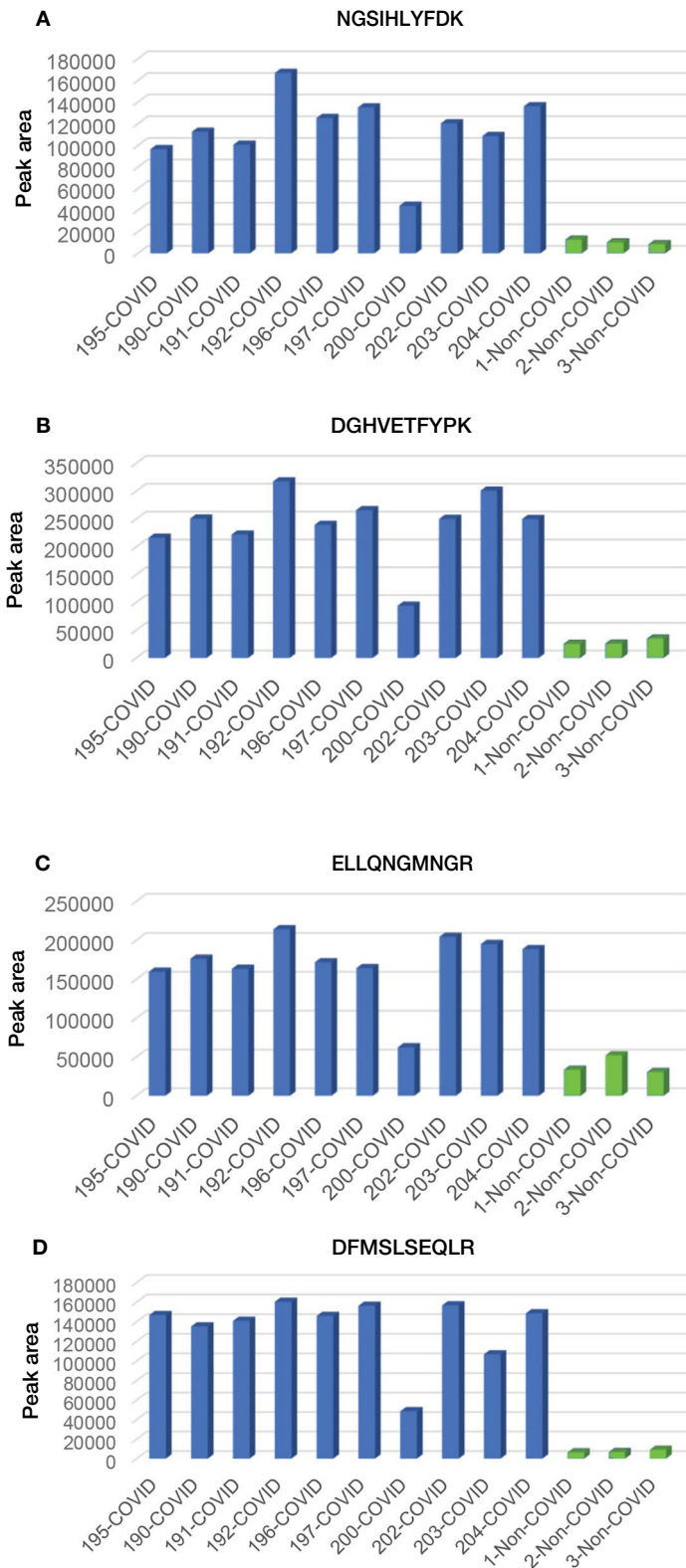
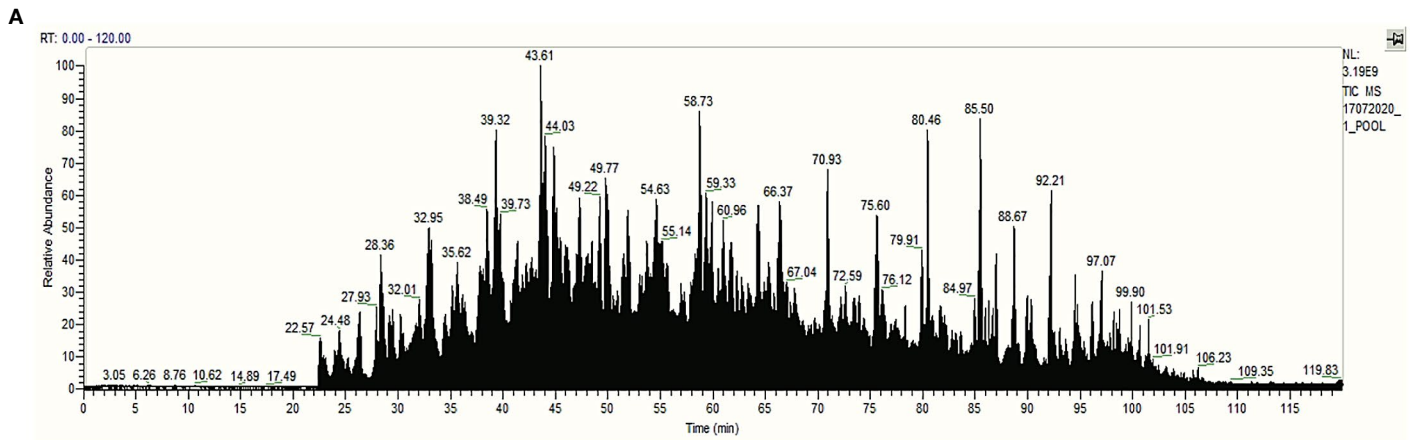
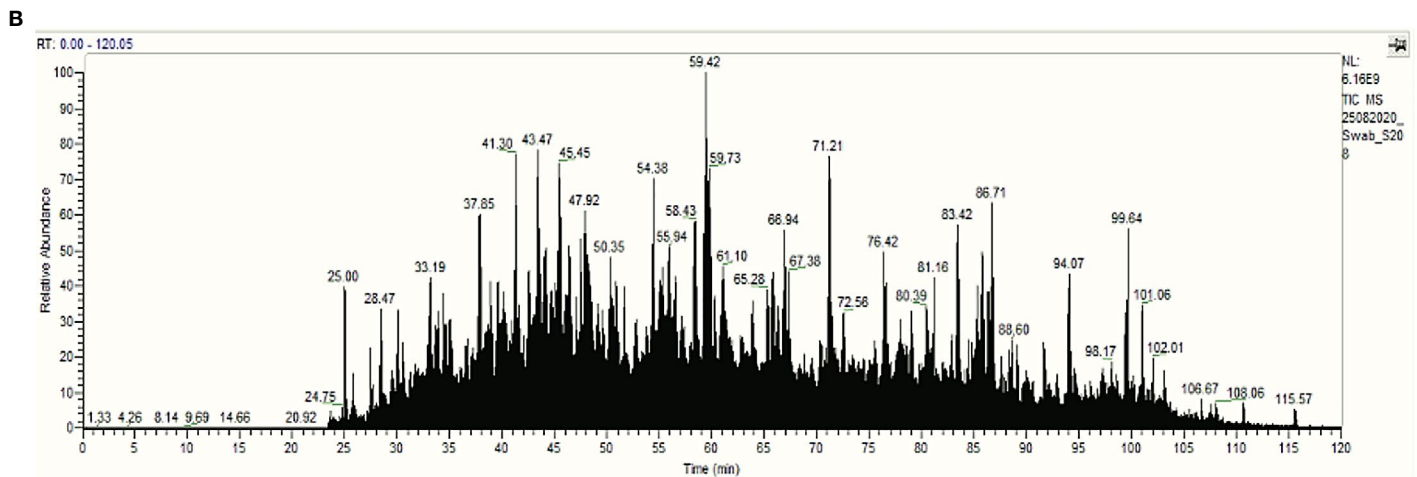


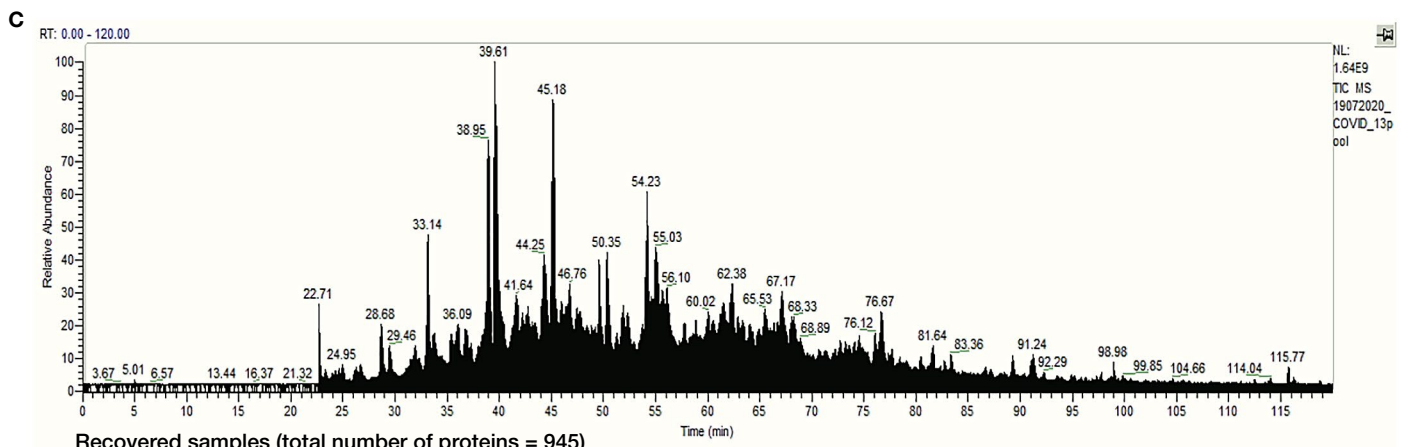
Figure 7. Peak area of peptides from SARS-CoV-2 replicase polyprotein 1ab detected in swab samples of COVID-19-infected individuals and healthy individuals (non-COVID). The peak areas were obtained from the MRM data using Skyline software.



Positive samples (total number of proteins = 873)



Negative samples (total number of proteins = 932)



Recovered samples (total number of proteins = 945)

Figure 8. Representative TIC chromatograms generated by mass spectrometry analysis of nasopharyngeal swab samples from COVID-19-infected positive (A), negative (B), and recovered samples (C)

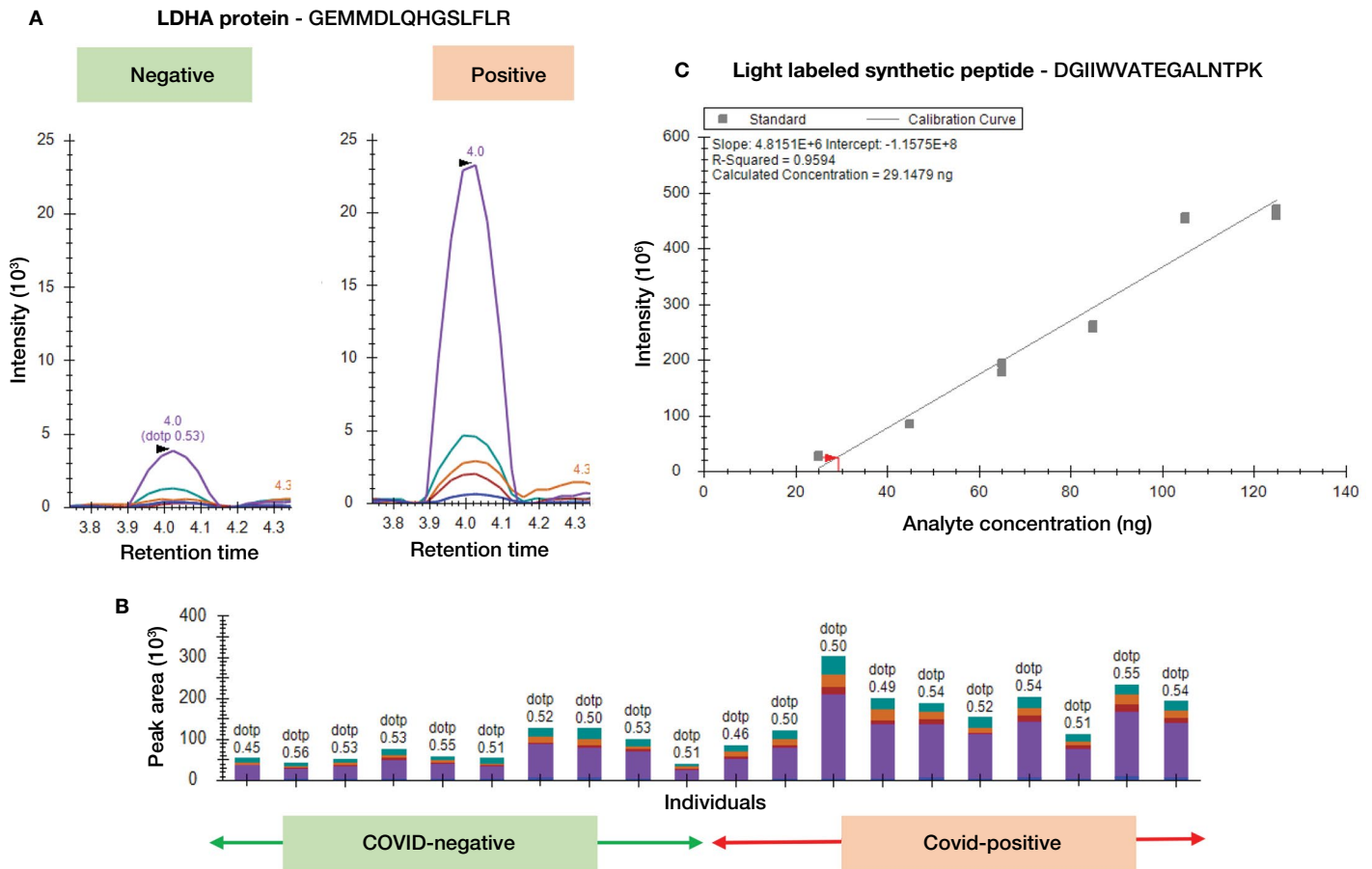


Figure 9. Targeted MRM-based validation of host peptide detected in the nasopharyngeal swab samples of COVID-19-infected samples. Intensity (A) and peak area (B) for the peptide specific for LDHA protein detected in COVID-19 negative and positive individuals; standard curve using MRM for light labeled synthetic peptide (C).

Discussion

A comprehensive understanding of the host response towards the viral infection can not only provide important insights regarding the progression of the disease from non-severe to severe but can also help in differentiating false positives and false negatives. The major goal of this study was to perform deep profiling of plasma and swab proteome from a cohort of COVID-19-infected individuals, facilitating the robust and statistically significant evaluation of differential expression between non-severe and severe disease groups.

In this study, several peptides belonging to SARS-CoV-2 nucleoprotein (P0DTC9), spike glycoprotein (P0DTC2), and replicase polyprotein 1ab (P0DTD1) in COVID-19-infected individuals were identified. A comprehensive method of peptide enrichment was used in the sample preparation

process, which ensured identification of a high number of viral peptides from the swab samples. The targeted LC-MS assays were developed to be simple, rapid, sensitive, and specific. Several transitions from SARS-CoV-2 peptides using the TSQ Altis triple quadrupole mass spectrometer were monitored, and 11 stable peptides, which showed distinct elution profiles and eluted between 0.5 to 8 min, were selected for routine monitoring. The SRM-based assay could specifically detect spike glycoprotein, nucleoprotein, and replicate viral peptides in COVID-19-infected individuals using high flow liquid chromatography.

A mass spectrometry-based label-free quantitative (LFQ) proteomics approach was explored to study alteration in plasma and swab proteome in COVID-19-infected individuals to understand the disease progression. Of the 1,200 proteins detected in the donor plasma, 38 proteins

were differentially expressed between non-severe and severe groups. The host proteins, such as angiotensinogen (AGT), apolipoprotein B (ApoB), SERPINA3, SERPING1, and fibrinogen gamma chain (FGG) identified in LFQ analysis, were further validated using the targeted mass spectrometry assay. Interestingly, previous studies have indicated that the increasing level of FGG correlated with the high level of D-dimers observed in the COVID-19-infected individuals with the worst outcomes. Thus, we certainly foresee the potential for this panel of host prognosis markers for clinical translation.

This observation not only displayed the sensitivity of the technology, but also the benefits of the additional specificity accessible using LC-MS for the screening of both symptomatic and asymptomatic population.

Conclusion

The host response usually differs in the SARS-CoV-2 infected individuals based on the immune strength and co-morbidity. Most of the severely infected individuals usually develop an acute phase response and breathlessness leading to an increased chance of fatality. A deep proteome profiling of COVID-19-infected respiratory samples based on label-free quantification using a high-resolution mass spectrometer was conducted to not only detect viral peptides, but also to identify significant host proteins that could be potential prognostic biomarkers for disease severity. The observations from this study can help in identifying critical host proteins as an optimal indicator

to diagnose and monitor the severity of this infection. The proteomics strategy could be used by the clinicians to understand more about the severity of the disease, and also the individual's recovery, which is not discernible from RT-PCR data.

This technical note provides discovery to quantitation workflows amongst COVID-19-infected individuals by monitoring two biological matrices – plasma and nasopharyngeal swab. The data presented here emphasizes that both NPS and plasma samples, which are routinely collected for the COVID-19 RT-PCR testing, can also be used for LC-MS based assay for detection of SARS-CoV-2 and assessment of the disease severity amongst individuals. The data suggests a list of promising biomarker candidates from respiratory and plasma samples. Additional tests to validate the findings from this study are in progress on a larger cohort with varying disease severity, along with blind analysis and simplified targeted SRM assays that can be easily implemented in every clinical research laboratory.

References

1. Tang, Y.; Schmitz, J.E.; Persing, D.H.; Stratton, C.W. Laboratory Diagnosis of COVID-19: Current Issues and Challenges. *J. Clin. Microbiol.* **2020**, *58*, 1–9. <https://jcm.asm.org/content/58/6/e00512-20>
2. COVID-19 plasma deep proteome reveals distinct signatures in severe patients. DOI: 10.21203/rs.3.rs-120570/v2 Preprint. <https://www.researchsquare.com/article/rs-120570/v2>
3. Proteomic investigation reveals dominant alterations of neutrophil degranulation and mRNA translation pathways in COVID-19 patients. <https://doi.org/10.1371/journal.pone.0240012>

Find out more at thermofisher.com/ClinicalApps