EBioMedicine 69 (2021) 103465

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

# A mass spectrometry-based targeted assay for detection of SARS-CoV-2 antigen from clinical specimens

Santosh Renuse[a,b,1], Patrick M. Vanderboom[a,1], Anthony D. Maus[a,1], Jennifer V. Kemp[a], Kari M. Gurtner[a], Anil K. Madugundu[a,c,d,e], Sandip Chavan[a], Jane A. Peterson[f], Benjamin J. Madden[f], Kiran K. Mangalaparthi[a,g], Dong-Gi Mun[a], Smrita Singh[a,c,d,e], Benjamin R. Kipp[h], Surendra Dasari[i], Ravinder J. Singh[a,*], Stefan K. Grebe[a,j,*], Akhilesh Pandey[a,b,e,*]

[a] Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, MN 55905, USA
[b] Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, USA
[c] Institute of Bioinformatics, International Technology Park, Bangalore, Karnataka 560066, India
[d] Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[e] Center for Molecular Medicine, National Institute of Mental Health and Neurosciences, Hosur Road, Bangalore, Karnataka 560029, India
[f] Proteomics Core, Medical Genome Facility, Mayo Clinic, Rochester, MN 55905, USA
[g] Amrita School of Biotechnology, Amrita Vishwa Vidyapeetham, Kollam, Kerala 690525, India
[h] Department of Laboratory Medicine and Pathology, Division of Laboratory Genetics and Genomics, Mayo Clinic, Rochester, MN 55905, USA
[i] Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA
[j] Department of Medicine, Division of Endocrinology, Mayo Clinic, Rochester, MN 55902, USA

ARTICLE INFO

*Article History:*
Received 12 March 2021
Revised 10 June 2021
Accepted 11 June 2021
Available online xxx

*Keywords:*
SARS-CoV-2
COVID-19
Diagnostic assays
Mass spectrometry
Ion mobility
Machine learning

ABSTRACT

*Background:* The COVID-19 pandemic caused by severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) has overwhelmed health systems worldwide and highlighted limitations of diagnostic testing. Several types of diagnostic tests including RT-PCR-based assays and antigen detection by lateral flow assays, each with their own strengths and weaknesses, have been developed and deployed in a short time.
*Methods:* Here, we describe an immunoaffinity purification approach followed a by high resolution mass spectrometry-based targeted qualitative assay capable of detecting SARS-CoV-2 viral antigen from nasopharyngeal swab samples. Based on our discovery experiments using purified virus, recombinant viral protein and nasopharyngeal swab samples from COVID-19 positive patients, nucleocapsid protein was selected as a target antigen. We then developed an automated antibody capture-based workflow coupled to targeted high-field asymmetric waveform ion mobility spectrometry (FAIMS) - parallel reaction monitoring (PRM) assay on an Orbitrap Exploris 480 mass spectrometer. An ensemble machine learning-based model for determining COVID-19 positive samples was developed using fragment ion intensities from the PRM data.
*Findings:* The optimized targeted assay, which was used to analyze 88 positive and 88 negative nasopharyngeal swab samples for validation, resulted in 98% (95% CI = 0.922–0.997) (86/88) sensitivity and 100% (95% CI = 0.958–1.000) (88/88) specificity using RT-PCR-based molecular testing as the reference method.
*Interpretation:* Our results demonstrate that direct detection of infectious agents from clinical samples by tandem mass spectrometry-based assays have potential to be deployed as diagnostic assays in clinical laboratories, which has hitherto been limited to analysis of pure microbial cultures.


## 1. Introduction

Coronaviruses are RNA viruses and include common human coronaviruses, severe acute respiratory syndrome-CoV (SARS-CoV), Middle East respiratory syndrome-CoV (MERS-CoV) and SARS-CoV-2 [1]. Although the large majority of common cold infections caused by coronaviruses do not have any significant clinical sequelae, the more recently identified coronaviruses have caused major public

* Corresponding authors at: Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, 200 First ST SW, Rochester, MN 55905, USA.
*E-mail addresses:* singh.ravinder@mayo.edu (R.J. Singh), grebe.stefan@mayo.edu (S.K. Grebe), pandey.akhilesh@mayo.edu (A. Pandey).
[1] Shared first authors

https://doi.org/10.1016/j.ebiom.2021.103465

**Research in Context**

*Evidence before this study*

At the beginning of the COVID-19 pandemic, there was a significant disruption in the supply chain for procuring reagents required for RT-PCR-based molecular diagnostic testing. At the time, there were no available viral antigen-based molecular tests with high sensitivity and sensitivity that could be used as alternatives to RT-PCR-based diagnostic testing. Mass spectrometry-based assays for identification of microbial pathogens were limited to detection of microbes from pure cultures and not capable of detecting pathogens directly from clinical specimens such as nasopharyngeal swabs.

*Added value of this study*

Using a unique combination of automated capture using anti-nucleocapsid antibody, rapid gradient chromatography, latest generation of Orbitrap mass spectrometers, ion mobility (FAIMS) and the use of machine learning for the final qualitative assay, this mass spectrometry-based immunoassay for detection of SARS-CoV-2 viral antigen provides a high sensitivity and specificity of viral detection.

*Implications of all the available evidence*

The approach described in this study for development of highly sensitive assays for detecting SARS-CoV-2 directly from clinical specimens can be applied to detecting other pathogens in clinical microbiology laboratories.

health outbreaks [2,3]. SARS-CoV-2 is the cause of the current COVID-19 pandemic [4,5]. Patients with COVID-19 present with upper respiratory infections that can result in a number of complications particularly in patients of advanced age or with co-morbidities. Recent reports show that in many severely ill patients it progresses to a multi-system disorder involving blood vessels (abnormal blood clotting), heart (acute cardiac injury, myocarditis and cardiac arrhythmias), kidneys (acute renal injury), liver, gut and brain [6–15].

The vast majority of current laboratory testing for COVID-19 is based either on detection of viral antigens, nucleic acids or on detection of virus-specific antibodies in the sera of patients. Viral antigen detection is generally based on immunoassays while viral nucleic acid detection is performed by real time quantitative RT-PCR, droplet digital PCR (ddPCR) or, less commonly, by techniques such as isothermal amplification [16]. Immunoassays offer moderate to good detection sensitivity and relative rapid analytical turnaround times, but some viruses such as enteroviruses and rhinoviruses have extensive antigenic heterogeneity, which makes them less amenable to antigen detection methods. Of the molecular methods, RT-PCR and dd-PCR typically have superior detection sensitivity compared to immunoassays and less problems with viruses that show antigenic shifts, while similar turnaround times to immunoassays. Detection of virus-specific antibodies allows for diagnosis of recent symptomatic or even asymptomatic infections. IgM antibodies are the first antibody isotype that develop rapidly within days after infection, while IgG levels take a week or longer to rise. Overall, the biggest advantage of nucleic acid-based assays is that a new assay for a previously unknown virus can generally be developed and validated in a matter of a days or a few weeks after the virus is first cultured, whereas development time lines for immunoassays are substantially longer [17,18]. However, the increased demand for PCR testing worldwide has resulted in test shortages and the acute need for more available accurate diagnostic tests.

Detection of viral material using liquid chromatography-tandem mass spectrometry (LC-MS/MS) might also have sufficient sensitivity for viral antigen detection. Previous studies have suggested that detection of viral proteins in body fluids could be a rapid diagnostic method for severe acute respiratory syndrome (SARS) [19–21]. Moreover, in the nearly two decades since the initial SARS outbreak, tremendous improvements have occurred in liquid chromatography as well as mass spectrometry and their applications in both research and clinical laboratories. The current generation of LC-MS-based approaches can be used for direct identification of viral proteins at relatively low concentrations and might therefore be suitable for clinical diagnostic applications. For instance, nucleocapsid protein from SARS patients was abundant enough to be detected in clinical samples using an ELISA approach [20]. There are limited studies published recently on the detection of the viral proteins in biological samples by mass spectrometry [22,23] and further described in Table s1. Gouveia et al. shortlisted 14 peptides from N, S and M proteins of SARS-CoV-2 with potential for use in development of MS-based targeted assays based on shotgun proteomics data from SARS-CoV-2 infected Vero cells [24]. Nikolaev et al. studied epithelial scrapings of nasopharynx for detection of peptides from N protein by tandem mass spectrometry [25] while Singh et al. described the detection of two peptides from SARS-CoV-2 using naso-oropharyngeal swab samples by a multiple reaction monitoring approach [26]. Cardozo et al. described a Sera-mag bead-based strategy for enrichment of viral proteins followed by turbo-flow chromatography mass spectrometry analysis to detect N protein-dereived peptides in nasopharyngeal swab samples with 84% sensitivity and 97% specificity [27]. Gouveia et al. utilized nasopharyngeal swabs spiked with purified SARS-CoV-2 virus as a proof-of-principle study to develop MS method for the detection of two peptides (ADETQALPQR and GFYAQGSR) from N protein [28] whereas Saadi et al. studied nasopharyngeal swab samples stored in viral transport medium (VTM) and described an LC-MS/MS assay targeting nine peptides from three viral proteins (N, S and M protein) [29]. The sensitivity of the assay was found to be limited to detection at Ct value of $\leq$24 of RT-PCR test, which reflects samples with high viral loads. Ihling et al. showed detection of viral peptides in gargle solution collected from three COVID-19 positive individuals [30]. As indicated, most of these reports are proof-of-principle experiments showing feasibility of viral protein detection by mass spectrometry and do not involve stress testing in low viral load samples. A MALDI-MS platform-based study described the potential for COVID19 diagnosis based on host response without any direct evidence of presence of viral proteins [31]. This method may be promising although it does not rule out the possibility of presence of other nasopharyngeal viral infections such as influenza owing to false positive diagnoses and therefore might not be amenable to deployment in clinical laboratories.

Amid nationwide vaccination, hundreds of thousands of newly diagnosed cases are being reported which necessitates continuous testing for COVID-19. Also, newer variants of COVID-19 are being reported which required further development of RT-PCR assays specific to different variants. Mass spectrometry-based diagnostic approaches could supplement and/or complement nucleic acid-based viral antigen testing. Here, we describe anti-nucleocapsid protein antibody-based approach for enrichment of nucleocapsid protein from nasopharyngeal swab samples of COVID-19 patients for development and evaluation of a mass spectrometry-based targeted qualitative assay. We deployed a supervised machine learning approach using a training dataset consisting of 187 samples with subsequent validation on 176 samples, which provided 98% sensitivity with 100% specificity.

## 2. Methods

### 2.1. COVID-19 specimen collection and handling and ethics

All samples were collected after informed consent and approval by the institutional review board (ID: 20-005649). Samples were collected in PBS and tested for SARS-CoV-2 by RT-PCR test at Mayo Clinic and deidentified prior to analysis. The Ct values from the reference RT-PCR test were used to generate Ct deciles and for selection of 10% of the positive cohort from each of the deciles. Thus, the Ct distribution of the samples analyzed in this study reflects the true distribution in the population as measured by the reference diagnostic method (RT-PCR test).

### 2.2. Chemicals and reagents

The following reagents were deposited by the Centers for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate USA-WA1/2020, gamma-irradiated (Catalog# NR-52287). Tris (2-carboxyethyl)phosphine hydrochloride (TCEP) (Catalog#C4706) and iodoacetamide (Catalog# I1149) were purchased from Sigma (St. Louis, MO). Phosphate buffered saline (PBS) was purchased from Bio-Rad (Catalog# 1610780, Hercules, CA). Trifluoroacetic acid (TFA) was purchased from Thermo Fisher Scientific (Catalog# 85183, Waltham, MA). Zwittergent Z3-16 was purchased from CalBiochem (Catalog# CAS 2281-11-0, EMD Millipore, Billerica, MA). Dynabeads M-280 Streptavidin was purchased from Invitrogen (Catalog# 11205D, Carlsbad, CA). Trypsin/Lys-C Mix (Catalog# V5071) and Rapid Digestion kit (Catalog# VA1060) was purchased from Promega (Madison, WI). Recombinant SARS-CoV-2 nucleocapsid protein was purchased from ProSci (Catalog# 97-077, Fort Collins, CO). Formic acid was purchased from Thermo Scientific (Catalog# TS-28905, San Jose, CA). Acetonitrile was purchased from J.T.Baker (Catalog# 75-05-8, Phillipsburg, NJ)

### 2.3. Synthesis of isotopically labeled heavy peptide standards

All SARS-CoV-2-derived peptides were synthesized using standard FMOC chemistry on a Liberty Blue (CEM Corp. Matthews, NC) peptide synthesizer with methods suggested by the manufacturer. Starting resin was either arginine (pbf) $^{13}C_6$, $^{15}N_4$ 2Cl-trt resin (Catalog# SRPR-ARG-CN-PK) or lysine (boc) $^{13}C_6$ $^{15}N_2$ 2Cl-trt resin (Catalog# SRPR-LYS-CN-PK) (Cambridge Isotope Laboratories, Tewksbury, MA), depending on the sequence. A second stable isotope labeled amino acid was incorporated into the sequence using Fmoc-Leucine-OH $^{13}C_6$ $^{15}N$ (Catalog# 608092), Fmoc-Glycine-OH $^{13}C_2$ $^{15}N$ (Catalog# 489522), or Fmoc-Alanine-OH $^{13}C_3$ (Catalog# 489867) (IsoTec, Inc., Miamisburg, OH) as dictated by the sequence. Peptides were cleaved using the CEM Razor cleavage module heated to 42 °C for 40 min. Cleavage cocktail was trifluoroacetic acid, water (Catalog# 14-650-357, Fisher Scientific, Hanover Park, IL), triisopropyl silane (Catalog# 233781, Sigma Aldrich, St. Louis, MO) and 3,6-Dioxa-1,8-octanedithiol (92.5%/2.5%/2.5%/2.5% v/v/v/v) (Catalog#D264925G, Fisher Scientific, Hanover Park, IL). Peptides were precipitated in cold methyl-t-butyl ether, washed (Catalog# D264925G, Fisher Scientific, Hanover Park, IL) and dried for purification. Purification was achieved by reversed-phase HPLC using a Phenomenex Jupiter $C_{18}$ column, 250 mm x 21.2 mm (Catalog# 00G-4057-P0, Torrance, Canada), using a water/acetonitrile buffer system. Peptide purity and integrity were confirmed using an Agilent InfinityLab II LC/MSD (Agilent Technologies, Santa Clara, CA) system.

### 2.4. Processing of nasopharyngeal swab samples and in-solution trypsin digestion

Nasopharyngeal swab samples were collected in PBS and digested directly. 100 $\mu$l of sample was diluted three-fold with rapid digest buffer (Promega, Madison WI) and 1.5 $\mu$g of trypsin was added. The samples were placed on a thermomixer (Thermo Scientific, Waltham, MA) at 70 °C for 1 h with rotation at 1,150 rpm for trypsin digestion. To terminate the digestion, the samples were acidified to 0.2% TFA. The samples were desalted using $C_{18}$ solid phase extraction spin columns (Catalog# TT2C18, Glygen, Columbia, MD), dried down and reconstituted in 20 $\mu$l of 0.2% formic acid. Recombinant proteins and inactivated purified SARS-CoV-2 were digested using the method described above for the swab samples with the following modifications: protein content of the purified SARS-CoV-2 was estimated using a BCA assay (Thermo Scientific, Waltham, MA) and the samples were concurrently reduced using 5 mM TCEP and in-solution trypsin digested at a 1:10 enzyme to substrate ratio. Following digestion, alkylation was carried out using 5 mM IAA. Samples were incubated in dark for 30 min at room temperature prior to sample acidification and cleanup.

### 2.5. Untargeted LC-MS/MS experiments

LC-MS/MS analysis for untargeted discovery proteomics experiments was carried out using an Ultimate 3000 RSLCnano system (Thermo Scientific, San Jose, CA) connected to an Orbitrap Eclipse mass spectrometer (Thermo Scientific, San Jose, CA). The peptides were loaded onto a trap column (PepMap $C_{18}$ 2 cm $\times$ 100 $\mu$m, 100 Å) at a flow rate of 20 $\mu$l/min using 0.1% formic acid and separated on an analytical column (EasySpray 50 cm $\times$ 75 $\mu$m, $C_{18}$ 1.9 $\mu$m, 100 Å, Thermo Scientific, San Jose, CA) with a flow rate of 300 nl/min with a linear gradient of 5 to 40% solvent B (100% ACN, 0.1% formic acid) over a 40 min gradient. Both precursor and fragment ions were acquired in the Orbitrap mass analyzer. Precursor ions were acquired in *m/z* range of 350−1,700 with a resolution of 120,000 (at *m/z* 200). Precursor fragmentation was carried out using higher-energy collisional dissociation (HCD) method using normalized collision energy (NCE) of 28. The fragment ions were acquired at a resolution of 30,000 (at *m/z* 200). The scans were arranged in top-speed method with 3 s cycle time between MS and MS/MS. Ion transfer capillary voltage was maintained at 2.5 kV. For internal mass calibration, lock mass option was enabled with polysiloxane ion (*m/z*, 445.120025) from ambient air.

### 2.6. Mass spectrometry data analysis of untargeted LC-MS/MS data

The raw mass spectrometry data were searched using Andromeda in MaxQuant software suite (version 1.6.7.0) [32] against a combined protein database of SARS-CoV-2 proteins, SARS-CoV proteins, common coronaviruses (OC43, HKU1, NL63 and L229E) and UniProt human protein database, African green monkey (*Chlorocebus aethiops*) database (in case of irradiated virus MS data) including common MS contaminants. The search parameters included a maximum of two missed cleavages; carbamidomethylation at cysteine as a fixed modification for samples that were reduced and alkylated; N-terminal acetylation and oxidation at methionine as variable modifications. Precursor tolerance was set to 10 ppm and MS/MS tolerance to $\pm$0.02 Da. False discovery rate was set to 1% at the peptide-spectrum matches (PSMs), peptide and protein levels.

### 2.7. Anti-nucleocapsid antibody-based enrichment of nucleocapsid protein and in-solution trypsin digestion

In order to improve the sensitivity of the detection, we evaluated several antibodies as shown in Table s2. Briefly, antibody was biotinylated using biotinylation kit (ThermoFisher Scientific, San Jose, CA) as per manufacturer's instructions. Biotinylated antibody (1 $\mu$g) was coated on streptavidin MSIA tips (Catalog#991STR11, ThermoFisher Scientific, Tempe, AZ) in 0.1% BSA containing 1X PBS on the Versette automated liquid handler (ThermoFisher Scientific, San Jose, CA). Nasopharyngeal swab samples (750 $\mu$l) were mixed with zwitterion

Z3-16 at final concentration of 0.002% in 96 well plate and were inactivated at 70 °C for 30 min. Inactivated samples were subjected to enrichment using mass spectrometry immunoassay (MSIA)-based enrichment using biotinylated antibody, washed two times with 200 $\mu$l 1X PBS and eluted in 100 $\mu$l of 50% ACN/0.002% Z3-16 in 0.1% TFA. Sample eluent was mixed with 300 $\mu$l of rapid trypsin digestion buffer (Promega Corporation, Madison, WI) and subjected to in-solution trypsin digestion (Gold Trypsin, Promega Corporation, Madison, WI) at 70 °C for 1 h on a shaker incubator. The digest was acidified using TFA to a final concentration of 1% TFA. The acidified digests were spiked-in with synthetic isotope labelled heavy peptides as retention time monitoring standards and the samples were loaded on EvoTips as per manufacturer's instructions. Quality control samples included recombinant nucleocapsid protein spiked into pool of negative samples, negative pooled sample, positive pooled samples with Ct values of 25, 29 and 33 based on results of the molecular test. Briefly, the $C_{18}$ EvoTips were activated using 20 $\mu$l of 100% acetonitrile followed by equilibration with 20 $\mu$l of 0.1% formic acid in water. Activation and equilibration was carried out at 700 x g for 1 min. The sample was loaded at 500 x g for 5 min followed by washing using 0.1% formic acid once. At last the tips were loaded with 100 $\mu$l of 0.1% formic acid and processed for targeted analysis.

### 2.8. Targeted parallel reaction monitoring (PRM) analysis

Parallel reaction monitoring (PRM) analysis was performed on an Exploris 480 mass spectrometer equipped with FAIMS Pro ion source (Thermo Scientific, Bremen, Germany) and interfaced with a preformed gradient LC system (EvoSep One, EvoSep Inc., Odense, Denmark). FAIMS Compensation voltages were optimized for peptides from nucleocapsid protein; AYNVTQAFGR and QQTVTLLPAADLDDFSK. Individual positive and negative swab samples were processed for antibody capture and trypsin digestion as described above. Peptides were eluted at a flow rate of 2 $\mu$l / minute into a pre-formed gradient using the Evosep One LC system connected to the Exploris 480 mass spectrometer equipped with FAIMS Pro ion source. Peptide separation was carried out using a 4 cm analytical column (Dr. Maisch $C_{18}$AQ, 1.9 $\mu$m, 150 $\mu$m x 4 cm) (Catalog# EV1107, EvoSep Inc., Odense, Denmark) with a 5.6 min gradient. Data acquisition parameters included MS1 scan from 560 to 1,000 m/z at resolution of 60,000 followed by retention time scheduled PRM analysis of AYNVTQAFGR and QQTVTLLPAADLDDFSK peptides and corresponding double SIL heavy peptides. The PRM parameters included: Orbitrap resolution of 60,000, AGC target value of $5 \times 10^4$, injection time of 118 ms, isolation window of *m/z* 1 and HCD normalized collision energy of 27. Table 1 shows targeted inclusion list with retention time scheduled PRM scans for light and heavy versions of AYNVTQAFGR and QQTVTLLPAADLDDFSK with corresponding FAIMS compensation voltages (CV). Quality control samples were analyzed before and after analysis of each batch of positive/negative samples.

### 2.9. Data analysis of targeted parallel reaction monitoring (PRM) data

The PRM data were processed using the Skyline software package [33]. Peak integration of all plausible fragments of analytes was carried out. The fragment ion intensities were exported from skyline and (natural) log transformed. A supervised machine learning method was used to select the optimal fragments and determine their weights for maximizing the performance of the targeted mass spectrometry assay. All computations were performed in R (version 4.0.1). For this, we utilized an ensemble-based machine learning approach encoded in the Super Learner [34]. This method was configured to use a generalized linear model via penalized maximum likelihood (glmNET), generalized linear model (glm) and random forest model; all configured to use binomial distribution. A 10-fold stratified cross-validation strategy with a goal to maximize the AUC was instituted during the learning process. Super Learner computed an optimally weighted average of the three different training models using the goal of maximizing the AUC while minimizing the risk scores (i.e. measure of training errors) of individual models. Final weighted ensemble model that produced the maximum AUC using three different models, each with risk score of $<$10%, on the training data was locked for an independent validation.

### 2.10. Accuracy, precision, and lower limit of detection

To evaluate the intra-day and inter-day accuracy and precision of the procedure, each validation batch was prepared on separate days. Each batch included the pooled nasopharyngeal swab samples with low (Ct pool of 25), medium (Ct pool of 29) and high (Ct pool of 33) Ct values, pooled negative sample and recombinant nucleocapsid protein spiked into pooled negative sample. All the QC samples were spiked-in with double isotopically labelled heavy synthetic peptides, AYNVTQAFGR and QQTVTLLPAADLDDFSK. Response curves were carried out for inactivated virus tryptic digest and heavy synthetic peptides to determine limit of detection (LOD). The assay precision was calculated using analyte to heavy internal standards ratio by analysis of 12 each pooled nasopharyngeal swab samples with low and high Ct values spiked-in with internal standards.

### 2.11. Statistics

All statistical analyses were performed in R (version 3.6.3). Sensitivity, specificity and predictive values were calculated using the "epiR" package. Linear regression was utilized to evaluate the linear relationship between RT-PCR Ct values and log2 transformed peptide intensities.

### 2.12. Role of funders

The funders had no role in the study design, writing of the manuscript or the decision to submit it for publication.

## 3. Results

We first sought to analyze the SARS-CoV-2 viral proteome by analyzing inactivated SARS-CoV-2 virus and recombinant viral proteins to determine which proteins and peptides were detectable.

**Table 1**
Peptides from nucleocapsid protein used for targeted FAIMS-PRM analysis for the detection of SARS-CoV-2 in nasopharyngeal swab samples. The table includes peptide sequence, precursor *m/z*, precursor charge, retention time (start and end) in minutes and optimized compensation voltage (CV).

| Peptide | m/z | Charge State | Start (min) | End (min) | FAIMS Pro CV |
|---|---|---|---|---|---|
| AYNVTQAFGR | 563.7856 | 2 | 2.5 | 3.5 | -40 |
| AYNVTQA(+3Da)FGR(+10Da) (Heavy IS) | 570.2948 | 2 | 2.5 | 3.5 | -40 |
| QQTVTLLPAADLDDFSK | 931.4807 | 2 | 3.5 | 5.3 | -30 |
| QQTVTLLPAADL(+7Da)DDFSK(+8Da) (Heavy IS) | 938.9964 | 2 | 3.5 | 5.3 | -30 |

### 3.1. SARS-CoV-2 proteome

An overview of the genome organization of SARS-CoV-2 is presented in Fig. 1a. This virus is most closely related to SARS-CoV and to MERS. The common human coronaviruses (OC43, HKU1, NL63 and L229E) are less related to SARS-CoV-2 than to each other. A sequence alignment of nucleocapsid protein in these coronaviruses is presented in Fig. s1. A schematic of domain and functional organization of two of the major structural proteins of SARS-CoV-2, along with the location of the peptides identified in our studies is shown in Fig. 1b.

### 3.2. Nucleocapsid protein as a target viral antigen for assay development

To characterize the SARS-CoV-2 proteome, we analyzed purified gamma-irradiated SARS-CoV-2 virus samples (USA-WA1/2020 isolate) cultured in African green monkey (*Chlorocebus aethiops*) kidney cells (Vero E6). The inactivated virus samples were digested with trypsin and analyzed on a high resolution Orbitrap Eclipse mass spectrometer interfaced with an RSLCnano system using a 3 h gradient. Searches against a database that included proteins from African green monkey, humans and SARS-CoV-2 plus other coronaviruses (SARS-CoV, MERS and common human coronaviruses). We included proteins from other coronaviruses to avoid proteins mistakenly identified from common coronaviruses (especially from COVID-19 negative patient samples as these individuals showed symptoms similar to other viral infection such as common cold). This resulted in identification of 951 proteins from monkey, 99 proteins from human (likely because of lack of a complete protein database for monkey) and 5 proteins (50 peptides) from SARS-CoV-2 including nucleocapsid (22 peptides), spike glycoprotein (22 peptides), membrane protein (2 peptides), orf1a (1 peptide) and orf9b (2 peptides) (Fig. 2a) (Table s3).

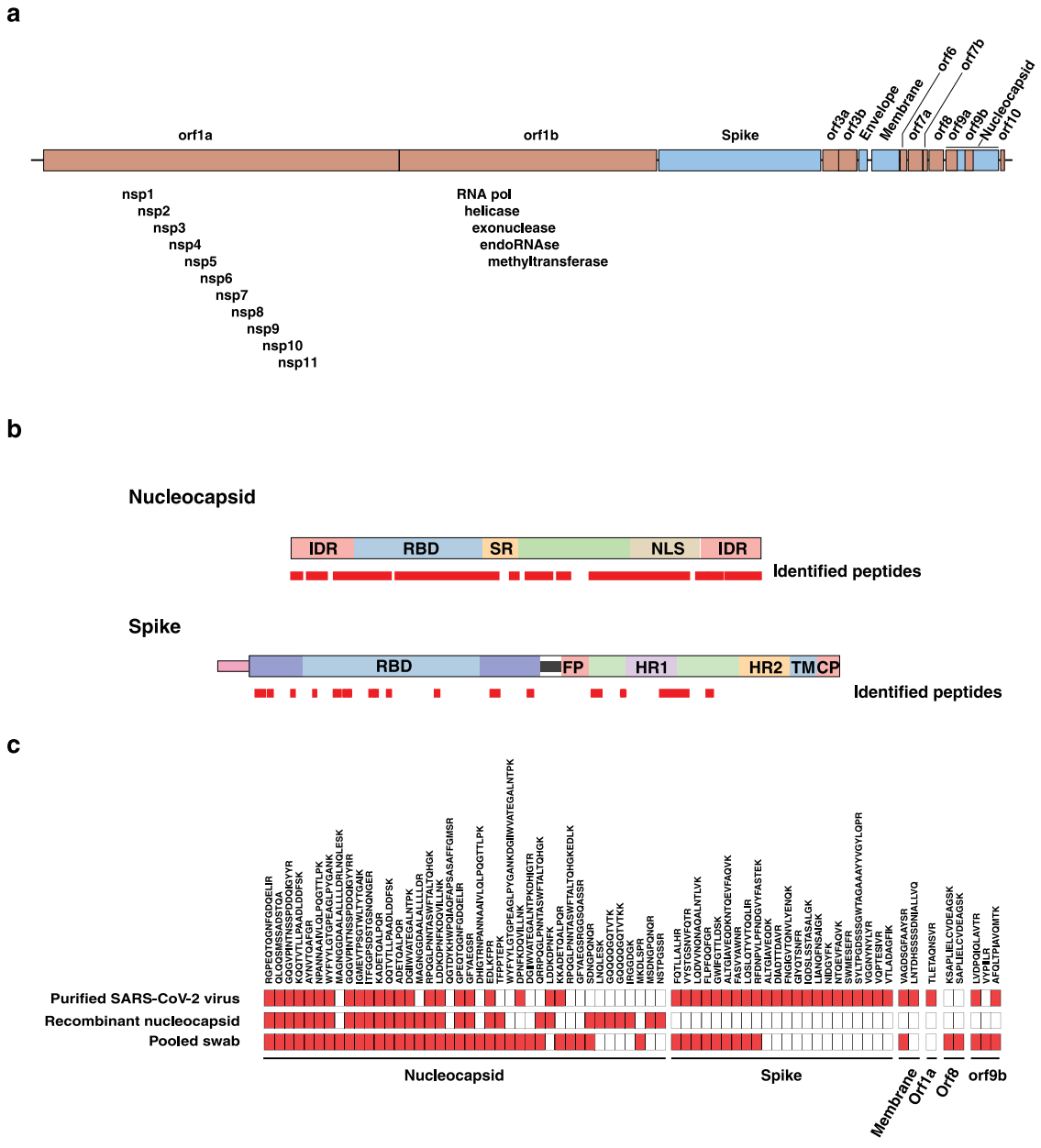


**Fig. 1.** Overview of the annotated genome (a) and domain organization of key structural components: nucleocapsid and spike proteins (b) of SARS-CoV-2 along with sequence coverage obtained by peptides identified by mass spectrometry in this study. Overview of proteins and peptides from SARS-CoV-2 identified in discovery studies by bottom-up mass spectrometry (c). Red boxes represent peptides that were detected while unfilled boxes represent peptides that were not detected in the indicated samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

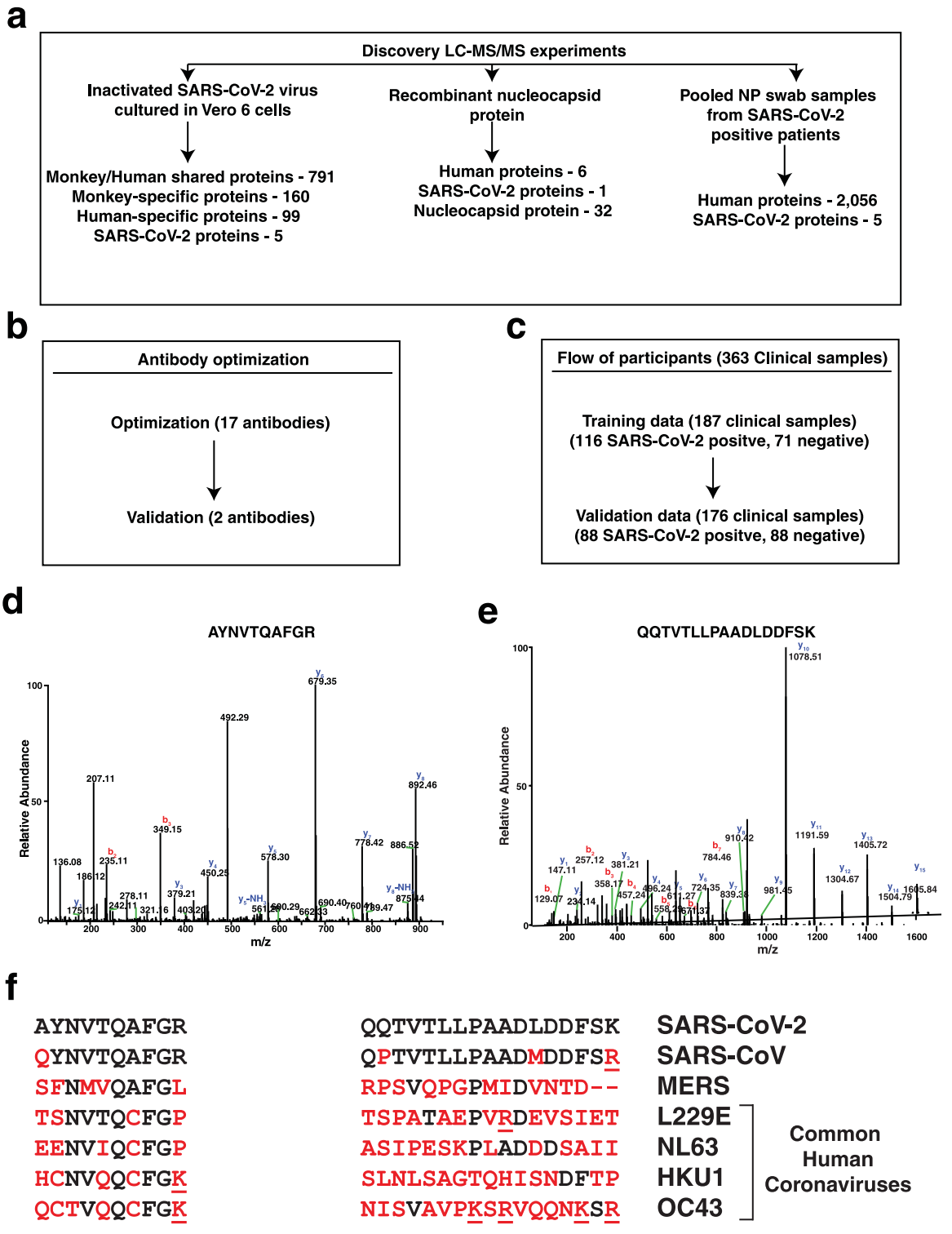


**Fig. 2.** A flowchart describing analysis flow such as discovery LC-MS/MS analysis indicating number of proteins/peptides identified in various samples (a); antibody optimization indicating shortlisting of 2 antibodies from total of 17 tested (b) and participant flow indicating total number of samples used for training (116 SARS-CoV-2 positive and 71 negative) and validation of the assay (88 SARS-CoV-2 positive and 88 negative) (c). Annotated MS/MS spectra of peptides identified from the nucleocapsid protein of SARS-CoV-2 positive nasopharyngeal swab specimens - AYNVTQAFGR (d) and QQTVTLLPAADLDDFSK (e). Sequence alignment of the two peptides that were chosen for development of PRM assays (f) - AYNVTQAFGR (267–276) and QQTVTLLPAADLDDFSK (389–405) - derived from SARS-CoV-2 nucleocapsid protein across related coronaviruses (SARS-CoV, MERS and common human coronaviruses: L229E, NL63, HKU1 and OC43). Amino acid mismatches are indicated in red while tryptic cleavage sites (R/K) are underlined. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Trypsin digestion and analysis by untargeted LC-MS/MS of recombinant nucleocapsid protein was undertaken to identify candidate peptides for targeted detection by LC-MS/MS assays. This analysis identified 32 peptides, most of which were unique to SARS-CoV-2 (13 were indistinguishable from SARS-CoV) (Fig. 2b) (Fig. s1). Peptides identified from SARS-CoV-2 recombinant nucleocapsid protein

are summarized in Fig. 1c and Table s2. To test if the peptides identified from these overexpression systems would also be identified from clinical samples, we performed LC-MS/MS-based deep proteomic profiling of pooled nasopharyngeal samples that had been confirmed by molecular test as SARS-CoV-2 positive. Database searches against human and SARS-CoV-2, SARS-CoV, MERS and other common human coronavirus proteins resulted in identification of 5,269 human proteins (50,540 peptides) and 5 proteins (55 peptides) from SARS-CoV-2 which included nucleocapsid protein (36 peptides), spike glycoprotein (3 peptides), membrane protein (1 peptide), orf9b (3 peptides) and orf8 (2 peptides), which was highly similar to the profiles previously obtained. Of the peptides that could uniquely identify SARS-CoV-2, two were chosen for targeted assay development for optimal detection (Fig. 2d,e) - all of them were unique to SARS-CoV-2 as shown by the sequence alignment with SARS-CoV, MERS and other common human coronaviruses (Fig. 2f).

### 3.3. Optimization of pre-analytical variables

Based on our discovery experiments, we decided upon nucleocapsid protein as our target antigen for assay development. We optimized several pre-analytical parameters including top performing peptides from nucleocapsid protein, protein extraction, various antibodies for capture of nucleocapsid protein and the use of ion mobility.

#### 3.3.1. Selection of top performing peptides from nucleocapsid protein

In the discovery analysis, we identified a total of 42 peptides from nucleocapsid protein from various sample types. We then shortlisted 10 peptides (Table s3) based on ionization efficiency, zero missed cleavages, limit of detection, carryover to finally narrow down to the best performing two peptides, AYNVTQAFGR and QQTVTLLPAADLDDFSK. Overall, these two peptides had significantly higher signal intensity under our sample preparation conditions than all other peptides that were evaluated. Subsequently, our assay was focused on these two peptides.

#### 3.3.2. Protein extraction optimization

To maximize protein yield from nasopharyngeal swab samples, we tested several protein extraction methods including methanol precipitation, 2% SDS, 0.002% Z3-16 zwitterionic detergent with and without incubation at 95 °C and RIPA buffer. The processed samples were immunopurified using anti-nucleocapsid protein antibody followed by targeted LC-MS/MS analysis on a Exploris 480 Orbitrap mass spectrometer coupled with EvoSep One fast flow liquid chromatography system. The parallel reaction monitoring (PRM) method included targeted analysis of AYNVTQAFGR and QQTVTLLPAADLDDFSK peptides from nucleocapsid protein. The data was analyzed using Skyline and the MS/MS fragment intensities for AYNVTQAFGR and QQTVTLLPAADLDDFSK peptides were compared across different protein extraction methods as shown in Fig. s2 and Table s4. We observed that 0.002% Z3-16 shown better extraction efficiencies as evident from fragment peak intensities and was selected for protein extraction and subsequent capture using antibody.

#### 3.3.3. Direct digest and immunopurification

Next, we tested if we could detect the nucleocapsid protein directly after proteolytic digest or if enrichment using an anti-nucleocapsid antibody prior to digestion was required for detection. For this comparison, a monoclonal antibody (Sino Biological, Wayne PA, Cat# 40143-R001) was biotinylated and immobilized to streptavidin mass spectrometry immunoassay (MSIA) tips for capture using an automated liquid handler. In all, 89 patient samples were analyzed with each method (69 positive and 20 negative as tested by RT-PCR). To determine if antibody enrichment resulted in an increase in sensitivity, a cutoff was established at 3 standard deviations above the average peak height observed for AYNVTQAFGR and QQTVTLLPAADLDDFSK peptides in the 20 RT-PCR negative samples. Peptide signals above the cutoff threshold in RT-PCR positive samples were then counted. Antibody enrichment resulted in 14 and 24 additional positive samples which were above the cutoff for AYNVTQAFGR and QQTVTLLPAADLDDFSK, respectively. The additional samples above the cutoff in antibody enriched samples had higher Ct values (i.e. lower viral loads) as compared to samples above the cutoff for both the methods, indicating that immunoaffinity enrichment results in a greater sensitivity relative to direct digest (Table s5).

#### 3.3.4. Antibody enrichment optimization

To improve the efficiency of nucleocapsid capture, 17 different antibodies against the nucleocapsid protein were evaluated (Table s2). Four antibodies which displayed the best performance characteristics were compared side by side using pooled patient samples with high, medium or low viral loads (based on Ct values from real time quantitative RT-PCR-based molecular testing). This comparison identified two monoclonal antibodies with suitable performance from which we selected one for further development of the assay.

### 3.4. Improved selectivity with ion mobility

In discovery LC-MS/MS analysis of eluates after nucleocapsid enrichment from nasopharyngeal samples, we detected >2,000 peptides. This poses unique challenges especially if short chromatographic separation is desired for targeted detection owing to interference from the highly complex matrix. Orthogonal gas phase separation such as ion mobility for separation of charged ions based on their size could potentially provide additional separation that might be beneficial. A front-end high-field asymmetric waveform ion mobility spectrometry (FAIMS) source typically filters ions entering into the ion transfer capillary by reducing background chemical noise resulting in increased selectivity and robustness, resulting in improved sensitivity [35,36]. To determine if FAIMS could increase selectivity for the viral peptides and improve the specificity of the method, we optimized compensation voltages for the peptides of interest from nucleocapsid protein. Those that resulted in the highest signal intensities were chosen to build the targeted method (Table s6). We analyzed nasopharyngeal swab samples with and without FAIMS and observed significantly improved signal-to-noise (S/N) ratio with FAIMS. Fig. s3a shows the PRM signal of AYNVTQAFGR peptide from a representative nasopharyngeal swab sample (relatively high viral load; Ct value of 20) where the S/N with FAIMS was 965 as compared to 67 without FAIMS. Fig. s3b shows the PRM signal for AYNVTQAFGR peptide in another representative nasopharyngeal swab sample with a low viral load (Ct value of 30) that could not be detected without FAIMS. Based on these data, we incorporated FAIMS into our assay workflow.

### 3.5. High-throughput assays for detection of SARS-CoV-2

To maintain optimal performance, the sample preparation and chromatography conditions were further refined. The sample preparation workflow was automated using Versette liquid handler which minimized overall time and experimental variation during sample processing. The throughput and robustness of the sample analysis was achieved by coupling Evosep One with Exploris 480 mass spectrometer, allowing analysis of 100 samples per day without sacrificing sensitivity.

Synthetic isotopically labeled heavy peptides were used for optimization of LC-MS/MS parameters including collision energy, retention time and relative intensity of fragment ions. As our aim was to develop a targeted qualitative assay, we used heavy peptide standards for retention time monitoring and as a control for manual peak integration to ensure accurate peak picking. Table 1 shows details of peptides including sequence, precursor m/z, charge, retention time

**Table 2**
Assay precision metrics showing CV values (%) calculated using analyte to IS ratios. Isotopically labelled heavy synthetic peptide standards were spiked into pooled nasopharyngeal swab samples with low and high Ct values and were analyzed by targeted LC-MS/MS assay for the detection of AYNVTQAFGR and QQTVTLLPAADLDDFSK peptides.

| Peptide Sequence | NP samples (Low Ct Pool) *n*=12 | | |
|---|---|---|---|
| | Average Analyte Height | Average internal STD Height | Analyte to IS ratio |
| AYNVTQAFGR | 455903 | 10214880 | 0.04 |
| CV (%) | 13.6 | 14.2 | 6.2 |
| QQTVTLLPAADLDDFSK | 156885 | 11385608 | 0.01 |
| CV (%) | 8.6 | 9.4 | 5.8 |
| | NP samples (High Ct Pool) n= 12 | | |
| AYNVTQAFGR | 3716094 | 6022087 | 0.6 |
| CV (%) | 26.6 | 26.8 | 6.5 |
| QQTVTLLPAADLDDFSK | 2323655 | 7520422 | 0.3 |
| CV (%) | 6.2 | 7.7 | 5.6 |

and FAIMS compensation voltages. Limit of detection (LOD) was determined for peptide AYNVTQAFGR using a double isotopically labelled heavy synthetic peptide and recombinant nucleocapsid protein. The LOD of the heavy synthetic peptide and recombinant nucleocapsid protein were observed to be 50 amol and 200 amol, respectively (Fig. s4a and b). Further, using gamma-irradiated SARS-CoV-2 virus (BEI Resources), we observed the LOD of the assay to be 2,000 genome equivalents of SARS-CoV-2 virions (Fig. s4c). The assay precision was calculated using isotopically labelled heavy peptide standard to analyte ratio by targeted LC-MS/MS analysis of 12 replicates of pooled nasopharyngeal swab samples with low and high Ct values spiked-in with internal standards and the coefficient of variation (CV) was observed to be <7% for both peptides (Table 2). This optimized qualitative assay was used to analyze 363 residual nasopharyngeal swab samples previously tested by quantitative real time RT-PCR. Quality control samples were included in each batch consisting of the pooled nasopharyngeal swab samples with low, medium and high Ct values, pooled negative sample and recombinant nucleocapsid protein spiked into pooled negative sample. Fig. 3 shows a schematic

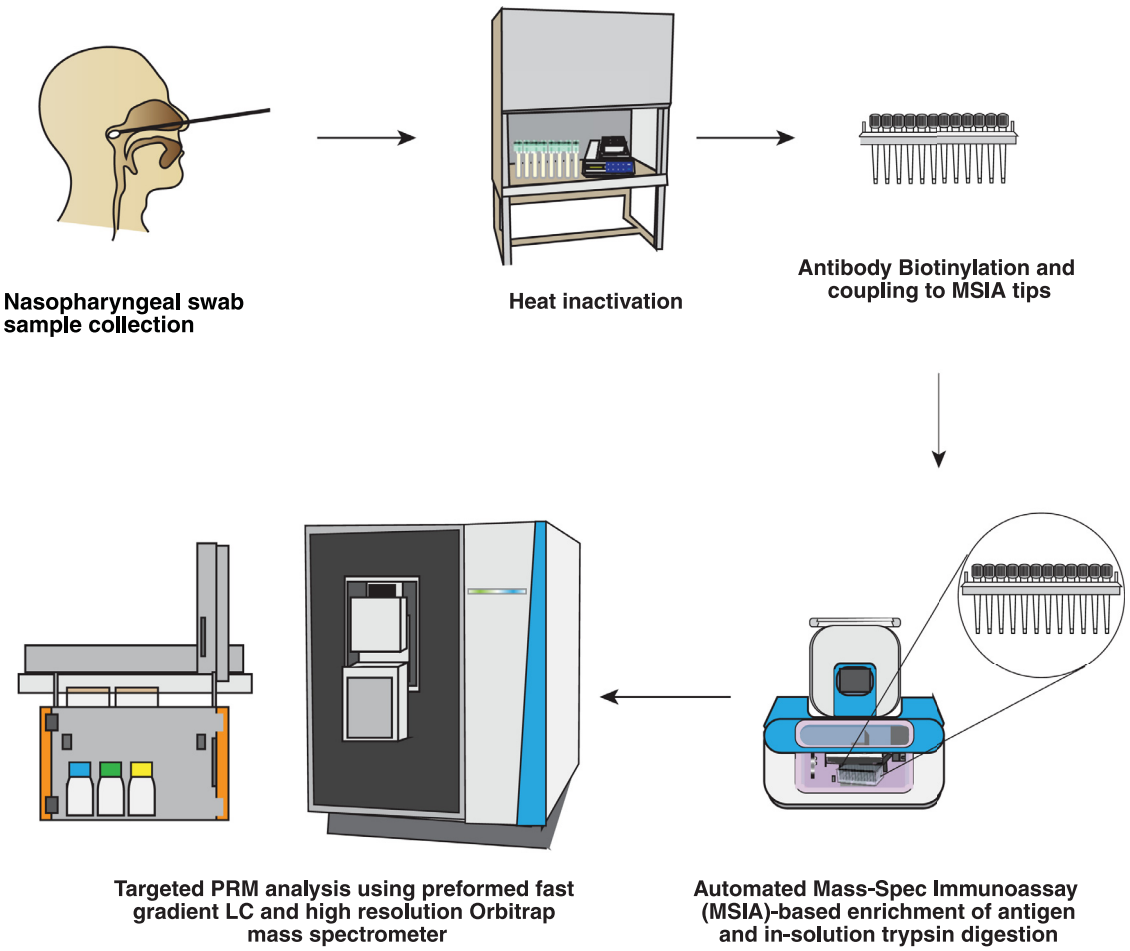


**Fig. 3.** A schematic of FAIMS-PRM targeted assay for the detection of diagnosis of SARS-CoV-2. Heat-inactivated nasopharyngeal swab samples were immunopurified with anti-nucleocapsid antibody coated on MSIA D.A.R.T.s tips using Versette automated liquid handler, in-solution trypsin digested. Targeted LC-MS/MS analysis was carried out on an Exploris 480 mass spectrometer interfaced with rapid chromatography system and a FAIMS Pro interface.

of optimized workflow for the diagnosis of SARS-CoV-2 using FAIMS-PRM targeted mass spectrometry.

We analyzed 187 nasopharyngeal swab samples (116 positive and 71 negative) using the optimized FAIMS-PRM workflow (Fig. 2c) and used supervised machine learning to select the optimal fragments and determine their weights for maximizing the performance of the targeted mass spectrometry assay. For this, we utilized Skyline software to manually integrate the fragment peaks (using heavy peptide as a marker) in all samples. All fragment intensities of light peptides were exported from skyline (without any peak filtering) and log transformed for statistical analysis ((Table s7). Using this as a training dataset, we employed an ensemble-based machine learning approach encoded in Super Learner as described previously [34]. This method was configured to use a generalized linear model via penalized maximum likelihood (glmNET), generalized linear model (GLM) and random forest (RF) model; all configured to use binomial distribution. A 10-fold cross-validation with a goal to maximize the AUC and to prevent overfitting was instituted during the learning process. An optimal weighted average of the trained models (i.e. glmNET, GLM and RF) was computed and considered as a final composite model to evaluate performance on an independent validation dataset. This ensemble method selected a total of 17 fragments (out of a total of 42) for incorporation into the model, resulting in an AUC of 0.9956 on the training set (Table 3). We also compared the performance of this machine learning model to a more traditional approach wherein the fragment ion intensities of each peptide were summed, and summed peptide intensities were utilized in a logistic model for discrimination. The AUC for this traditional method was 0.9117 (Fig. s5). We did not attempt to transform the summed peptide intensities into absolute quantification values for discrimination because this transformation does not substantially increase the information content of the data beyond the summed peptide intensity data and the machine learning method working with individual fragment data has already achieved near perfect discrimination. Thus, we chose the superior performing machine learning model for independent validation. Based on the probability distribution of negative and positive samples of the machine learning model in the training set, we chose a probability of $\geq$0.6 as threshold for calling a positive sample (Fig. 4a and b). A probabilistic threshold of 0.6 was chosen using the training dataset that maximized positive predictive value (PPV). This value was chosen before validation samples were processed for mass spectrometry and data analysis We applied this composite model to an independent validation dataset of 176 samples (88 positive and 88 negative samples) using the exact same optimized FAIMS-PRM workflow and fragment ion data processing (Table s7). In order to avoid any undue influence of training on validation, samples in the validation data set were acquired and processed after the machine learning model was trained on the training data set and locked for validation. The model achieved a sensitivity of 98% (95% CI = 0.922–0.997) and specificity of 100% (95% CI = 0.958–1.000) on the independent validation data set (Table 3). These results show that our assay for the detection of SARS-CoV-2 antigen performed with a very high sensitivity and specificity, especially compared to other reported rapid antigen lateral flow assays [37].

### 3.6. Assay performance as compared to RT-PCR tests

Although RT-PCR remains the well-established gold standard for viral pathogen detection, the FAIMS-PRM workflow displays very high sensitivity and specificity. To ensure that samples used for training and validation of the model reflect the distribution observed in a large population, we mimicked the distribution of Ct values for 11,575 SARS-CoV-2 positive patients tested at Mayo Clinic (Fig. 4c) in the training and test set samples (Fig. 4d). Ct values obtained from the large population approximate a normal distribution (mean = 26.5, IQR = 21.1 – 31.3) and agree well with the training and validation samples (mean = 26.5, IQR = 21.8 –31.4). While naïve, uncalibrated Ct values are not themselves quantitative they can serve as an estimate of relative target abundance. Peptide peaks heights derived from mass spectrometry analysis also provide semi quantitative information. To determine the feasibility of using summed fragment intensities (peak heights) as an estimate of viral abundance, Ct values were correlated to peak heights using the training and validation data sets (Fig. 4e and f) and Table s8. Indeed, both peptides display good correlation with Ct values (QQTVTLLPAADLDDFSK $R^2$ = 0.768 (95% CI = 0.706–0.819), AYNVTQAFGR $R^2$ = 0.755 (95% CI = 0.690–0.808)) indicating that viral peptide peak height may additionally serve as an estimate of viral abundance. In a separate set of COVID-19 positive samples ($n$ = 44) (Table s9), we carried out dd-PCR-based viral copy number estimation. As shown in Fig. s6a, the exponential nature of the crossing thresholds (Ct) of the RT-PCR necessitates a non-linear fit and there is a reasonable agreement between dd-PCR-based copy numbers and the Ct values $\leq$30. However, at Ct values greater than this, the viral copy numbers and Ct values do not correlate well. Additionally, a distribution of predicted probability against RT-PCR Ct values for validation dataset containing 88 positive samples is shown in Fig. s6b. As shown by dotted line in Fig. s6b, 2 out of 88 samples were below the predicted probability of 0.6.

## 4. Discussion

There are limited published reports on SARS-CoV-2 viral proteome [24,35,36] or recombinant viral proteins [36]. Basu et al. discussed the merits of mass spectrometry-based proteomics SARS-CoV-2 research and testing considering data from recent publications or manuscript on preprint servers [37]. We and others have identified several peptides from SARS-CoV-2 either from inactivated virus cultured in mammalian cells or recombinant viral proteins. Based on our discovery data, the next step was to select the most promising peptides from SARS-CoV-2 based on spectral abundance and response intensity to create targeted PRM methods to characterize their performance. Finally, we developed targeted assays deploying automated antibody capture-based workflow followed by a rapid separation low-flow LC method and a PRM-based method incorporating ion mobility for selected viral peptides. The overall analysis scheme is shown in Fig. s7. In all, we tested 363 nasopharyngeal residual swab samples from patients with matched clinical molecular test results.

SARS-CoV-2 nucleocapsid is highly expressed during infection and a potential target for viral detection by mass spectrometry. The nucleocapsid protein forms a supercoiled helix structure and helps

**Table 3**
Sensitivity, specificity and predictive values for the training and validation data. 95% confidence intervals are indicated in parenthesis.

| Machine Learning: Training Data | | | | |
|---|---|---|---|---|
| AUC | PPV (%) | NPV (%) | Sensitivity (%) | Specificity (%) |
| 0.996 (98.9–1.00) | 98.6 (0.924–1.00) | 93.1 (0.869–0.970) | 89.7 (0.808–0.955) | 99.1 (0.951–1.00) |
| **Machine Learning: Validation Data** | | | | |
| AUC | PPV (%) | NPV (%) | Sensitivity (%) | Specificity (%) |
| 1.00 (1.0 – 1.0) | 100 (0.959, 1.00) | 97.7 (0.920–0.997) | 97.8 (0.922–0.997) | 100 (0.958–1.000) |

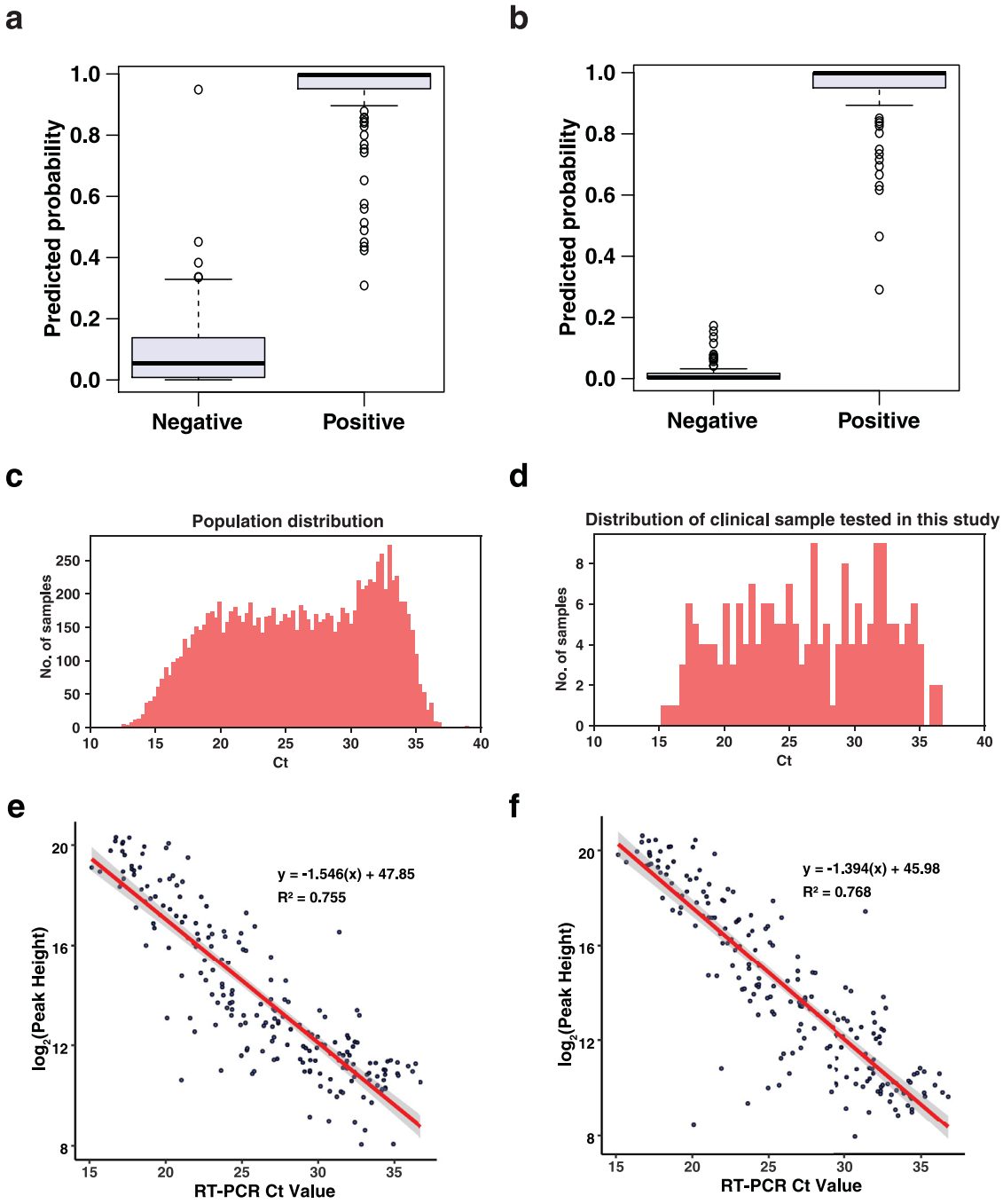


**Fig. 4.** Box plot showing predicted probability for training dataset (116 SARS-CoV-2 positive and 71 negative samples) (a) and validation dataset (88 SARS-CoV-2 positive and 88 negative samples) (b). Population distribution of nasopharyngeal swab samples tested by RT-PCR assay at Mayo Clinic containing (c), Distribution of clinical samples tested using targeted FAIMS-PRM method in this study (88 SARS-CoV-2 positive samples) (d), Targeted FAIMS-PRM $log_2$ transformed summed fragment ion intensities against RT-PCR Ct values of nasopharyngeal swab samples for AYNVTQAFGR (e) and QQTVTLLPAADLDDFSK (f) peptides from 88 SARS-CoV-2 positive samples.

package the single stranded viral RNA genome with at least 1,000 nucleocapsid molecules per virion [38,39]. A cell infected with SARS-CoV-2 might contain several virions and even larger numbers will be produced before budding of new virus commences; thus, there is a possibility to detect even relatively low viral loads in patient samples with modern LC-MS/MS instrumentation. Indeed, we consistently detected nucleocapsid-derived peptides as the most abundant peptides in our studies of the inactivated virus. Gouveia et al. [24] described seven peptides from nucleocapsid protein using inactivated SARS-CoV-2 virus cultured in Vero cells and all of them were identified in our study.

In this study, we report development of a mass spectrometry-based targeted assay using microflow chromatography that permits direct detection of viral antigens from clinical specimens. A key advantage of this targeted method is the high specificity and comparable sensitivity to the gold standard RT-PCR method for the diagnosis of SARS-CoV-2. Preliminary discovery mass spectrometry analysis of SARS-CoV-2 positive samples led to identification of suitable candidate peptides from nucleocapsid protein. We started with 10 most abundantly identified peptides from nucleocapsid protein that could serve as proteotypic peptides. We tested various antibodies against nucleocapsid protein for optimization of efficient capture from

clinical specimens. Our final assay employs a monoclonal antibody against nucleocapsid protein which is a perpetual source. Notably, the assay described in this study shows significantly better sensitivity and specificity as compared to recently reported targeted mass spectrometry assays.

We observed potential carryover after analysis of samples with high viral load (CT <24) even though we used EvoTips which are discarded. The carryover was found to be due to analytical column. We modified LC method with additional high organic washes at the end of the gradient and included at least three method blank analyses after every sample. This is one of the important considerations in analysis of viral infected samples in large scale testing which could have a significant impact on false positives. Although, this resulted in a drop in throughput of our assay, but it guarantees a high specificity by eliminating carryover. One of the limitations of our current assay is sample throughput which is currently 100 samples/day. We expect that established clinical laboratories with Orbitrap instrumentation are capable of deploying this assay quite easily and overall cost of the assay expected to be comparable or only slightly higher than that most RT-PCR or POC tests. Finally, the assay described in this study requires well-trained personnel and mass spectrometrists, which is an important consideration for the successful implementation of the assay.

With the introduction of several vaccines, we anticipate that the number of new infections will slowly decrease. However, it would take at least 6–12 months to vaccinate entire US population and even more so worldwide. Newer variants are being reported which further complicates accurate detection by existing molecular testing and requires development of new molecular assays. Mass spectrometry assays require little to no development for detection of new variants based on the location of mutation of spike or nucleocapsid protein and could be rapidly developed if any of the major mutations affected the peptides currently being detected or if the goal was to detect specific variants.

Definitive identification by this method can also serve as a reference method for other antigen testing assays such lateral flow assays, which have lower sensitivity [40]. In addition, the method can be applied to other sample types (e.g. urine, saliva) and can be adapted for detection of novel variants of SARS-CoV-2 as they appear. Finally, our mass spectrometry-based method is highly sensitive, utilizes rapid chromatographic separation for a throughput of 100 samples per day, and has the potential to be deployed at routine clinical laboratories with appropriate mass spectrometry instrumentation. Prior to consideration for clinical use, however, it is imperative to establish assay performance metrics by performing a rigorous validation using appropriate standards and controls.

## Declaration of Competing Interest

All authors declared no conflict of interest.

## Data sharing statement

The mass spectrometry raw data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD020644. Skyline files have been deposited to Panorama Public and can be accessed using unique identifier URL (https://panoramaweb.org/covid19ml.url).

## Contributors

A.P., S.K.G., R.J.S., B.R.K., S.R., P.M.V., A.D.M designed the study; A. P., S.R., P.M.V. managed the study; S.R., P.M.V., A.D.M, A.P. developed the pipeline of experiment and analysis; A.P., S.R., P.M.V., S.S., D.G.M. performed literature search; S.R., P.M.V, T.D.M., J.V.K., K.M.V., D.G.M., S.C., B.J.M., K.K.M. processed and prepared the samples for LC-MS/MS; J.A.P. performed peptide synthesis; S.R., P.M.V., T.D.M. performed LC-MS/MS; T.D.M., P.M.V., S.D., S.R., A.K.M. processed MS data, performed bioinformatic analysis; S.R., P.M.V., T.D.M., A.K.M., D.G.M., prepared figures and tables with help of other authors; S.R., P.M.V., T. D.M., A.K.M., S.D., A.P. verified the underlaying data; S.R., P.M.V., T.D. M., A.P. wrote the manuscript with help of other authors. All authors read and approved the manuscript.

## Acknowledgments

The following reagent was deposited by the Centers for Disease Control and Prevention and obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate USA-WA1/2020, Gamma-Irradiated, NR-52287. This study was supported by DBT/Wellcome Trust India Alliance Margdarshi Fellowship grant IA/M/15/1/502023 awarded to AP and the generosity of Eric and Wendy Schmidt.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103465.

## References

[1] Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 2019;17(3):181–92.
[2] Zhong NS, Zheng BJ, Li YM, Poon XZH, Chan KH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet 2003;362(9393):1353–8.
[3] Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 2012;367(19):1814–20.
[4] Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020;395(10223):470–3.
[5] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020;382 (13):1199–207.
[6] Madjid M, Safavi-Naeini P, Solomon SD, Vardeny O. Potential effects of coronaviruses on the cardiovascular system: a review. JAMA Cardiol 2020.
[7] Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020.
[8] Cheng Y, Luo R, Wang K, Zhang M, Wang Z, Dong L, et al. Kidney disease is associated with in-hospital death of patients with COVID-19. Kidney Int 2020;97 (5):829–38.
[9] Wong SH, Lui RN, Sung JJ. Covid-19 and the digestive system. J Gastroenterol Hepatol 2020;35(5):744–8. doi: 10.1111/jgh.15047.
[10] Lin L, Jiang X, Zhang Z, Huang S, Zhang Z, Fang Z, et al. Gastrointestinal symptoms of 95 cases with SARS-CoV-2 infection. Gut 2020;69(9):997–1001. doi: 10.1136/gutjnl-2020-321013.
[11] Li LY, Wu W, Chen S, Gu JW, Li XL, Song HJ, et al. Digestive system involvement of novel coronavirus infection: prevention and control infection from a gastroenterology perspective. J Dig Dis 2020;21(4):199–204. doi: 10.1111/1751-2980.12862.
[12] Zhang C, Shi L, Wang FS. Liver injury in COVID-19: management and challenges. Lancet Gastroenterol Hepatol 2020;5(5):428–30.
[13] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan. China. Lancet. 2020;395 (10223):497–506.
[14] Musa S. Hepatic and gastrointestinal involvement in coronavirus disease 2019 (COVID-19): what do we know till now? Arab J Gastroenterol 2020;21(1):3–8.
[15] Li YC, Bai WZ, Hashikawa T. The neuroinvasive potential of SARS-CoV2 may play a role in the respiratory failure of COVID-19 patients. J Med Virol 2020;92(6):552–5. doi: 10.1002/jmv.25728.
[16] Udugama B, Kadhiresan P, Kozlowski HN, Malekjahani A, Osborne M, Li VYC, et al. Diagnosing COVID-19: the disease and tools for detection. ACS Nano 2020;14 (4):3822–35.
[17] Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. N Engl J Med 2020;382 (12):1177–9.
[18] Hong KH, Lee SW, Kim TS, Huh HJ, Lee J, Kim SY, et al. Guidelines for laboratory diagnosis of coronavirus disease 2019 (COVID-19) in Korea. Ann Lab Med 2020;40(5):351–60.
[19] Kammila S, Das D, Bhatnagar PK, Sunwoo HH, Zayas-Zamora G, King M, et al. A rapid point of care immunoswab assay for SARS-CoV detection. J Virol Methods 2008;152(1-2):77–84.
[20] Che XY, Hao W, Wang Y, Di B, Yin K, Xu YC, et al. Nucleocapsid protein as early diagnostic marker for SARS. Emerg Infect Dis 2004;10(11):1947–9.

[21] Cho SJ, Woo HM, Kim KS, Oh JW, Jeong YJ. Novel system for detecting SARS coronavirus nucleocapsid protein using an ssDNA aptamer. J Biosci Bioeng 2011;112(6):535–40.
[22] Grossegesse M, Hartkopf F, Nitsche A, Schaade L, Doellinger J, Muth T. Perspective on proteomics for virus detection in clinical samples. J Proteome Res 2020;19(11):4380–8.
[23] Grenga L, Armengaud J. Proteomics in the COVID-19 battlefield: first semester check-up. Proteomics 2021;21(1):e2000198.
[24] Gouveia D, Grenga L, Gaillard JC, Gallais F, Bellanger L, Pible O, et al. Shortlisting SARS-CoV-2 peptides for targeted studies from experimental data-dependent acquisition tandem mass spectrometry data. Proteomics 2020;20(14):e2000107.
[25] Nikolaev EN, Indeykina MI, Brzhozovskiy AG, Bugrova AE, Kononikhin AS, Starodubtseva NL, et al. Mass-spectrometric detection of SARS-CoV-2 virus in scrapings of the epithelium of the nasopharynx of infected patients via nucleocapsid N protein. J Proteome Res 2020;19(11):4393–7. doi: 10.1021/acs.jproteome.0c00412.
[26] Singh P, Chakraborty R, Marwal R, Radhakrishan VS, Bhaskar AK, Vashisht H, et al. A rapid and sensitive method to detect SARS-CoV-2 virus using targeted-mass spectrometry. J Proteins Proteom 2020;11:159–65. doi: 10.1007/s42485-020-00044-9.
[27] Cardozo KHM, Lebkuchen A, Okai GG, Schuch RA, Viana LG, Olive AN, 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.
[28] Gouveia D, Miotello G, Gallais F, Gaillard JC, Debroas S, Bellanger L, et al. Proteotyping SARS-CoV-2 virus from nasopharyngeal swabs: a proof-of-concept focused on a 3 min mass spectrometry window. J Proteome Res 2020;19(11):4407–16. doi: 10.1021/acs.jproteome.0c00535.
[29] Saadi J, Oueslati S, Bellanger L, Gallais F, Dortet L, Roque-Afonso AM, et al. Quantitative assessment of SARS-CoV-2 virus in nasopharyngeal swabs stored in transport medium by a straightforward LC-MS/MS assay targeting nucleocapsid, membrane, and spike proteins. J Proteome Res 2021;20(2):1434–43.
[30] Ihling C, Tanzler D, Hagemann S, Kehlen A, Huttelmaier S, Arlt C, et al. Mass spectrometric identification of SARS-CoV-2 proteins from gargle solution samples of COVID-19 patients. J Proteome Res 2020;19(11):4389–92. doi: 10.1021/acs.jproteome.0c00280.
[31] Nachtigall FM, Pereira A, Trofymchuk OS, Santos LS. Detection of SARS-CoV-2 in nasal swabs using MALDI-MS. Nat Biotechnol 2020;38(10):1168–73. doi: 10.1038/s41587-020-0644-7.
[32] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008;26(12):1367–72.
[33] MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 2010;26(7):966–8.
[34] van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6:25.
[35] Swearingen KE, Moritz RL. High-field asymmetric waveform ion mobility spectrometry for mass spectrometry-based proteomics. Expert Rev Proteom 2012;9(5):505–17.
[36] Bekker-Jensen DB, Martinez-Val A, Steigerwald S, Ruther P, Fort KL, Arrey TN, et al. A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. Mol Cell Proteom 2020;19(4):716–29.
[37] Basu A, Zinger T, Inglima K, Woo KM, Atie O, Yurasits L, et al. Performance of Abbott ID NOW COVID-19 rapid nucleic acid amplification test in nasopharyngeal swabs transported in viral media and dry nasal swabs, in a New York City academic institution. J Clin Microbiol 2020;58(8):1–7. doi: 10.1128/JCM.01136-20.
[38] Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein–forms and functions. Antiviral Res 2014;103:39–50.
[39] Bar-On YM, Flamholz A, Phillips R, Milo R. SARS-CoV-2 (COVID-19) by the numbers. Elife 2020;9(e57309). doi: 10.7554/eLife.57309.
[40] Dinnes J, Deeks JJ, Adriano A, Berhane S, Davenport C, Dittrich S, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. Cochrane Database Syst Rev 2020;8:CD013705.