ORIGINAL ARTICLE

A rapid and sensitive method to detect SARS-CoV-2 virus using targeted-mass spectrometry

Praveen Singh[1,2] · Rahul Chakraborty[1,2] · Robin Marwal[3] · V. S. Radhakrishan[3] · Akash Kumar Bhaskar[1,2] · Himanshu Vashisht[3] · Mahesh S. Dhar[3] · Shalini Pradhan[1] · Gyan Ranjan[1,2] · Mohamed Imran[1,2] · Anurag Raj[1,2] · Uma Sharma[3] · Priyanka Singh[3] · Hemlata Lall[3] · Meena Dutta[3] · Parth Garg[4] · Arjun Ray[4] · Debasis Dash[1,2] · Sridhar Sivasubbu[1,2] · Hema Gogia[3] · Preeti Madan[3] · Sandhya Kabra[3] · Sujeet K. Singh[3] · Anurag Agrawal[1,2] · Partha Rakshit[3] · Pramod Kumar[3] · Shantanu Sengupta[1,2]

Abstract

In the last few months, there has been a global catastrophic outbreak of severe acute respiratory syndrome disease caused by the novel coronavirus SARS-CoV-2 affecting millions of people worldwide. Early diagnosis and isolation are key to contain the rapid spread of the virus. Towards this goal, we report a simple, sensitive and rapid method to detect the virus using a targeted mass spectrometric approach, which can directly detect the presence of virus from naso-oropharyngeal swabs. Using a multiple reaction monitoring we can detect the presence of two peptides specific to SARS-CoV-2 in a 2.3 min gradient run with 100% specificity and 90.5% sensitivity when compared to RT-PCR. Importantly, we further show that these peptides could be detected even in the patients who have recovered from the symptoms and have tested negative for the virus by RT-PCR highlighting the sensitivity of the technique. This method has the translational potential of in terms of the rapid diagnostics of symptomatic and asymptomatic COVID-19 and can augment current methods available for diagnosis of SARS-CoV-2.

Keywords SARS-CoV-2 · COVID-19 · LC-MS/MS · Proteomics · MRM · Naso-oropharyngeal swab

---

Praveen Singh, Rahul Chakraborty, Robin Marwal, Radhakrishan have contributed equally.

✉ Partha Rakshit
  partho_rakshit@yahoo.com

✉ Pramod Kumar
  pramodjnu1@gmail.com

✉ Shantanu Sengupta
  shantanus@igib.res.in

[1] CSIR-Institute of Genomics and Integrative Biology, Mathura Road, New Delhi 110025, India

[2] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

[3] National Center for Disease Control, New Delhi 110054, India

[4] Department of Computational Biology, Indraprastha Institute of Information Technology, Okhla Industrial Estate, Phase III, New Delhi 110020, India

Introduction

The world is in the midst of a pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). More than 15 million individuals from 200 countries have been infected with this virus, which resulted in over 600,000 deaths (https://www.worldometers.info/coronavirus). Disconcertingly, even after seven months from the first reported case in Wuhan, China, in December 2019 (Wang et al. 2020), the number of cases are on the rise with more than 200,000 new cases being reported daily (https://www.worldometers.info/coronavirus). One of the ways to reduce disease propagation is early diagnosis and isolating people infected with the virus. For this, mass testing is the need of the hour. At such a scale, it is critical to innovatively leverage existing infrastructure to build surge capacity and investigate low cost, high throughput methods.

Among the various methods available or developed recently, for diagnosis of SARS-CoV-2, RT-PCR is still the gold standard. Despite RT-PCR being well established and

widely available, many problems have emerged in actual performance, ranging from sub-optimal sensitivity, especially as the virus mutates, to low throughput due to many intermediate steps and long reaction times(Tahamtan and Ardebili 2020). Additionally, specialized reagents are necessary for RT-PCR, which increases the cost of the test. Apart from this, rapid antigen tests are also being extensively used in various countries for preliminary screening. These tests are based on the detection of antigens from the nasopharyngeal swab using antibodies. A positive test confirms the presence of the virus, but a negative test is inconclusive since the sensitivity of the antigen test is between 34 and 80% (Bruning et al. 2017). Thus, a highly sensitive method is required to directly detect the virus from the nasopharyngeal swab to facilitate rapid screening. Mass spectrometer with high accuracy and sensitivity can be an ideal platform to identify viral peptides with medium to high throughput.

Mass spectrometry (MS) based viral peptide detection has earlier been used for detection of viral proteins (Majchrzykiewicz-Koehorst et al. 2015; Foster et al. 2015; Santana et al. 2014). Identification of SARS-CoV-2 specific peptides have also been reported recently (Gouveia et al. 2020; IIhling et al. 2020). Gouveia et al. used SARS-CoV-2 infected cell lysate to identify peptides and validated these peptides in nasopharyngeal swab (Gouveia et al. 2020). They also shortlisted several peptides for targeted mass spectrometric studies (Gouveia et al. 2020). Cardozo et al. have reported the detection of three peptides using a 10.5 min run time with 83% sensitivity and 96% specificity (Cardozo et al. 2020). Ihling et al. used a gargle solution of 3 infected patients and report the identification of 1 peptide from nucleoprotein using a 180 min gradient (IIhling et al. 2020).

In this study, we report a simple and rapid method that could detect two peptides, QIAPGQTGK and AIVSTIQRKYK, from structural spike glycoprotein and replicase polyprotein 1ab respectively in 2.3 min gradient with a sensitivity of 90% and specificity of 100%. Using follow up samples of patients who have symptomatically recovered and also tested negative for RT-PCR analyses, we show that these peptides are present in these samples indicating the potential of asymptomatic diagnosis of COVID-19 in patients.

Methods

Clinical sample collection

Upper respiratory tract sample (nasopharyngeal swab and oropharyngeal swab) were collected in virus transport medium and samples were analyzed using COBAS 6800 automated system (Roche, Basel)according to the manufacturers guidelines(Corman et al. 2020). This study was

approved by institutional ethics committees of National Centre for Disease Control, New Delhi and CSIR-Institute of Genomics and Integrative Biology, Delhi.

Sample preparation

Naso and oropharyngeal swab collected in viral transport media (VTM) were inactivated by incubation of samples with lysis buffer (25% guanidinium thiocyanate and 5% SDS) for 20 min at room temperature (Pastorino et al. 2020). To this, 100 µl sodium deoxycholate (0.15% w/v) per ml of sample was added and was incubated for 10 min at room temperature followed by addition of 20% (w/v) trichloroacetic acid (TCA) for protein precipitation. The samples were then centrifuged at 15,000 g for 15 min at 4 °C. Pellets were washed three times with pre-chilled acetone and protein pellets were resuspended in 100 mM Tris–HCl with 8 M urea (pH-8.5).

Identification of SARS-CoV-2 proteins using high-resolution mass spectrometry

Proteins extracted from the nasopharyngeal swab of eight RT-PCR confirmed COVID-19 patients were pooled and 300 µg of protein was reduced with 25 mM dithiothreitol for 25 min at 56 °C, followed by alkylation using 55 mM iodoacetamide (IAA) at room temperature for 20 min in dark. This sample was subjected to trypsin digestion (sequencing grade, Promega) using an enzyme–substrate ratio of 1:10 for 18 h at 37 °C and dried under vacuum. The digested peptides were then fractionated into 8 fractions using cation exchange chromatography using a SCX Cartridge (5 micron, 300 Å bead from AB Sciex, USA).A step-gradient of increasing concentration of ammonium formate buffer (35 mM, 50 mM, 75 mM, 100 mM, 125 mM, 150 mM, 250 mM and 350 mM ammonium formate, 30% v/v acetonitrile and 0.1% formic acid; pH=2.9) was used to elute the peptides.

LC–MS/MS data acquisition

The fractionated peptides were analyzed on a quadrupole-TOF hybrid mass spectrometer (TripleTOF 6600, Sciex, USA) coupled to a nano-LC system (Eksigent NanoLC-425) in data-dependent acquisition (DDA) mode. Peptides from each fractions were cleaned using C18 Ziptip (Merck, USA) and 4 µg of these peptides were loaded on a trap-column (ChromXP C18CL 5 µm 120 Å, Eksigent) where desalting was performed with a flow rate of 10 µl per minutes for 10 min. Peptides were separated on a reverse-phase C18 analytical column (ChromXP C18, 3 µm 120 Å, Eksigent) at a flow rate of 5 µl/minute using buffer A (water with 0.1% formic acid) and buffer B (acetonitrile with 0.1% formic

acid) with following gradient: buffer B was increased gradually from 3 to 25% in first 38 min. It was increased to 32% solvent B in the next 5 min, in the next 2 min buffer B was ramped up to 80% and 90% in further 0.5 min. It was held at the same concentration for the next 2.5 min. Buffer B was brought to the initial 3% concentration in the next 1 min and column was reconditioned for 8 min before next run.

Ion source parameters were set to 5.5 kV for the ion spray voltage, 25 psi for the curtain gas, 15 psi for the nebulizer gas and 250 °C as temperature. For DDA, a 1.8 s instrument cycle was repeated in high sensitivity mode throughout the whole gradient, consisting of a full scan MS spectrum (400–1250 m/z) with an accumulation time of 0.25 s, followed by 30 MS/MS experiments (100–1500 m/z) with 0.050 s accumulation time each, on MS precursors with charge state 2 + to 5 + exceeding a 150 cps threshold. Rolling collision energy was used and former target ions were excluded for 15 s.

Database search

For viral protein identification, a merged search for 8 DDA runs was performed in Protein Pilot software v5.0.1 (Sciex, USA) with paragon algorithm. The parameters used were as follows: Cysteine alkylation—IAA, digestion—trypsin and 2 missed cleavages were allowed. The search effort was set to 'Thorough ID' and false discovery rate (FDR) analysis was enabled. Proteins identified with 1% global FDR were considered. The search was carried out against a UniProt database containing thirteen SARS-CoV-2 proteins.

Selection of peptides for validation phase

Protein BLAST search

Protein BLAST search was performed for each of the identified peptide with non-redundant protein sequences (nr) database using NCBI blastp suite using default parameters for short sequence search.

Peptide variability analysis

To check whether the identified peptides fall in the invariable region of proteins in SARS-CoV-2, their sequence homology was checked with protein sequences of around 54,000 SARS-CoV-2 strains submitted in the Global Initiative on Sharing All Influenza Data (GISAID) in FASTA format (version 01July2020). Different protein sequences were segregated and aligned, to observe the variability of the identified peptides amongst different strains. JalView tool was used for viewing the consensus of peptides over all the strains (Waterhouse et al. 2009).

Detection of SARS-CoV-2 specific proteins

Method standardization

A short scheduled multiple reaction monitoring (sMRM) method, with a retention time window of 20 s was developed with the unique viral peptides identified in DDA runs, with multiple transitions for each of the peptides. Sample preparation was performed as described above and data acquisition was performed using a triple quadrupole hybrid ion trap mass spectrometer (QTrap 6500 +, Sciex) coupled with a Sciex ExionLC UHPLC system. Tryptic peptides were loaded on an Acquity CSH C18 column (1.7 µm 2.1 × 100 mm, Waters) and separated using a binary gradient with buffer A (0.1% formic acid in water) and buffer B (0.1% formic acid in acetonitrile). Peptides were loaded on column with a flow-rate of 600 µL per minute and 98% buffer A. Buffer B concentration was increased from an initial 2–10% in 0.2 min and 50% in 1.1minutes. Buffer B concentration was ramped-up to 90% in another 0.2 min and held at the same gradient for 0.1 min and then brought to initial 2% in the next 0.1minute and held at the same concentration for 0.6. The total gradient was for 2.3 min. Peak review was performed using MultiQuant 3.0 (Sciex).

Serology test for detection of antibody against SARS-CoV-2

To check if the individuals had developed antibodies against SARS-CoV-2, we collected blood samples (5 ml) in vacutainer tubes from individuals and separated the plasma by centrifugation. Presence of antibodies was detected using Elecsys anti-SARS-CoV-2 on Cobas e411, Roche.

Results

Identification of SARS-CoV-2 proteins using high-resolution mass spectrometry

We initially performed a data-dependent acquisition of a pool of eight symptomatic RT-PCR confirmed patients using high-resolution mass spectrometer as mentioned in the methods section. We identified 22 peptides from 4 proteins of SARS-CoV-2with 1% False discovery rate (FDR). We identified three structural proteins-spike glycoprotein (spike), nucleoprotein (NP) and envelope small membrane protein (ENV) and a protein Replicase polyprotein 1ab (Rep) from non-structural part (Supplementary Table 1)-Among these, eight peptides from 3 proteins (Rep, Spike and NP), that were unique and un-modified was selected for generation of

Table 1 List of unshared and un-modified peptides identified in database search for data dependent run through high-resolution mass spectrometry (HRMS)

UniProtKB accession no.	Protein name	Peptide sequence	Charge	Parent m/z
P0DTD1	Replicase polyprotein 1ab	AIVSTIQRKYK	3	436.2662
		LTDNVYIK	2	483.2687
		MDGSIIQFPN	2	561.2684
P0DTC2	Spike glycoprotein	LIANQFNSAIGK	2	638.3564
		STNLVKNK	2	452.2665
		AHFPREGVFVS-NGTHWFVTQR	4	618.8134
		QIAPGQTGK	2	450.2508
P0DTC9	Nucleoprotein	ADETQALPQR	2	564.7858

an in-house sMRM method for SARS-CoV-2 specific protein detection in patients (Table 1).

Protein BLAST search

We also performed Protein BLAST searches with non-redundant protein sequences (nr) database using NCBI blast suite. The search revealed that seven of the eight peptides were unique to SARS-CoV-2. They were AIVSTIQRKYK and MDGSIIQFPN of Rep, QIAPGQTGK, LIANQFNSAIGK, AHFPREGVFVSNGTHWFVTQR and STNLVKNK of Spike and ADETQALPQR of NP.

Peptides in conserved regions

Sequences for more than 54,000 strains of SARS-CoV-2 have been documented in GISAID up to July 1, 2020. This gave us an opportunity to check if the peptides selected were in the invariant region. Screening through protein sequences of all these strains we found these peptides to be completely conserved among different SARS-CoV2 strains and showed no variability in the peptide sequences (Supplementary Fig. 1). Few strain sequences in the GISAID database show gaps and variations, probably due to low sequence quality or host editing.

Detection of SARS-CoV-2 specific proteins

For diagnostic purposes, it is important to obtain the desired results in the shortest possible time. Keeping this in mind, we developed a short scheduled multiple reaction monitoring (sMRM) method of 2.3 min with multiple transitions for each of the identified peptides. Based on the consistency and quality of the peaks in the ion chromatogram we included two peptides- AIVSTIQRKYK from Rep and QIAPGQTGK from spike in the sMRM method (Fig. 1). Apart from that, we also included an additional peptide of serum albumin as an internal control for protein amount (Fig. 1). The optimized parameters used for

the MRM method, to detect SARS-CoV-2 proteins in patient samples are provided in Supplementary Table 2. We considered a sample to be positive for SARS-CoV-2 if overlapping transitions were present with an intensity of greater than 1.5×10^3 and a signal to noise ratio of greater than 10 for either of the two selected peptides.

Detection of SARS-CoV-2 in naso-oropharyngeal swab samples

We analyzed 103 samples using the sMRM method developed in two different sample sets. The first set consisted of 20 samples (including follow up) from 14 patients for whom we had to follow-up RT-PCR data (Table 2). Interestingly, although most of the patients had recovered as was evident from their being RT-PCR negative and being free of symptoms, we found that they continue to retain the peptides even after 14 days (Sample 5, 6, 7, 8, 10 in Table 2) of initial infection. In 2 of the patients (sample 1 and 14), RT-PCR was found to be negative while the patients had symptoms. One of these patients (sample 14) was found to be negative on the 1st day but was found positive on the second day while we found peptide peaks to be present on both the days, clearly indicating the sensitivity of the sMRM method.

In the second set, we analyzed 82 case control samples. Among these 63 were RT-PCR positive and considered as patients. For controls, we considered 20 individuals who had negative RT-PCR and did not have any antibodies as evident from negative serology test indicating the absence of IgM and IgG against SARS-CoV-2. Mass spectrometric analysis revealed that in all the 20 negative samples, peptides were not detected. Of the 63 positive samples, 57 were found to be positive using our method. Thus, our method has a sensitivity of 90.4% and specificity of 100% with respect to RT-PCR positive samples and true controls (serologically negative, RT-PCR negative).

Fig. 1 a Chromatographic separation of the three peptides of naso-oropharyngeal protein digest. Chromatogram for the selected fragment ions are shown in di_erent colours for tryptic digest peptides **b** (i) AIVSTIQRKYK (Replicase polyprotein 1 ab), (ii) QIAPGQTGK (Spike glycoprotein) and (iii) AEFAEVSK (Albumin)
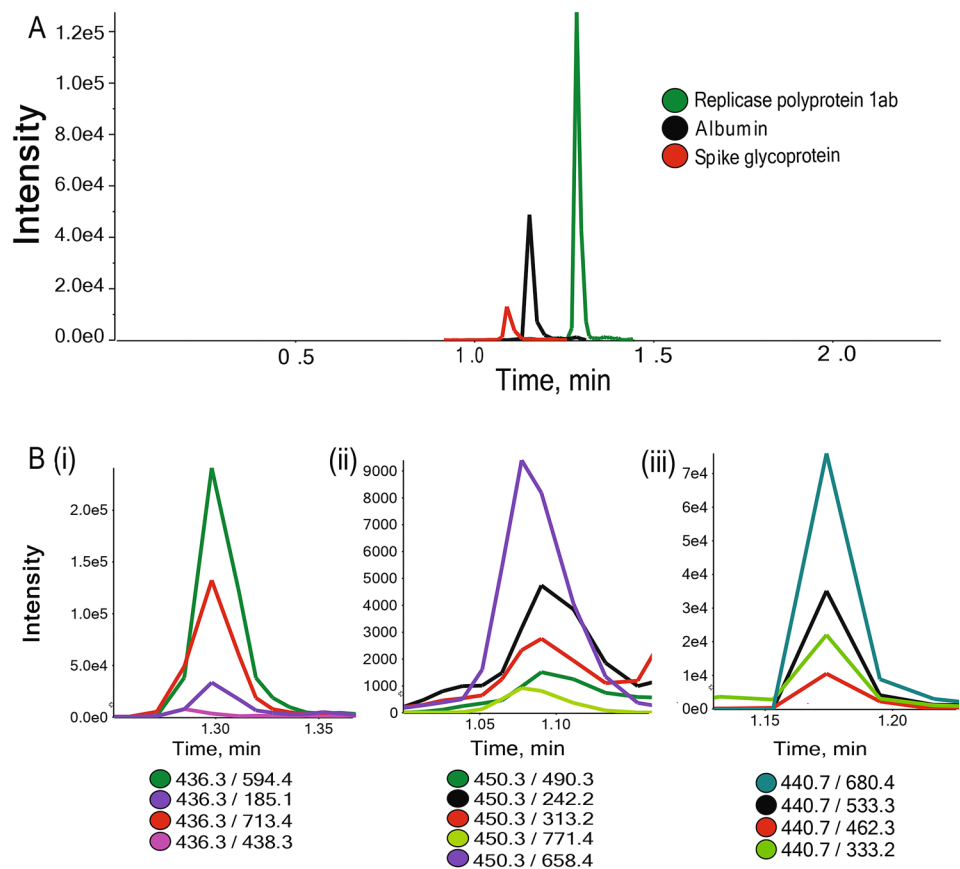
Table 2 List of patients considered for set-1 analysis

Index patient	Date after 1st RTPCR	Symptoms	RT PCR	Mass spec
1	2	Yes	Negative	Positive
2	2	No	Negative	Positive
3	3	Yes	Positive	Positive
4	13	Yes	Positive	Positive
5	0	Yes	Positive	Positive
	10	Yes	Positive	Positive
	18	No	Negative	Positive
6	12	No	Negative	Positive
	16	No	Negative	Positive
7	14	No	Negative	Positive
8	12	Yes	Positive	Positive
	17	No	Negative	Positive
9	10	No	Negative	Positive
10	10	No	Negative	Positive
	14	No	Negative	Positive
11	10	No	Negative	Positive
12	5	No	Negative	Positive
13	1	No	Negative	Positive
14	0	Yes	Negative	Positive
	1	Yes	Positive	Positive

Discussion

The objective of the study was to develop a simple and rapid targeted mass spectrometric method (sMRM) to identify peptides specific to SARS-CoV-2. For this, we initially analyzed samples in high-resolution mass spectrometer and identified eight unique and unmodified peptides from three proteins of SARS-CoV-2. These peptides were unique to SARS-CoV-2 and most importantly lay in the invariant region as revealed by analysis of more than 54,000 SARS-CoV-2 strains sequences submitted in GSAID up to 1st July 2020. This indicates that even if there are mutations in the virus, it should not affect the sequence of the peptides chosen. Using these peptides, we developed a short, 2.3 min, the gradient for the sMRM method and selected two peptides AIVSTIQRKYK from Replicase and QIAPGQTGK from Spike protein of SARS CoV-2 based on the consistency and quality of the spectra. Apart from these two peptides, we also included one peptide of human serum albumin AEFAEVSK as a control for protein amount. This was important since in a few samples we found that the intensities of albumin transitions were very low indicating low protein concentration and if the viral load is low in such samples it could result in false-negative results. These samples were thus not considered.

Using this method we analyzed 103 naso-oropharyngeal samples in two sets. In the first set, we included 20 samples from 14 patients for which RT-PCR data of two or more time points was available. Interestingly, peptide peaks were detected in patients who recovered from the symptoms and were RT-PCR negative for SARS-CoV-2. Interestingly, peptide peaks were observed in two symptomatic patients who tested negative with RT-PCR of which one was tested positive in subsequent RT-PCR indicating false negative RT-PCR in the first instance. Thus, the data indicates high sensitivity of the current method and can be used for initial diagnosis including diagnosis of asymptomatic cases.

In the second set, we analyzed case–control samples. As discussed previously, patients who have recovered, and were RT-PCR negative also show peptide peaks. For this, we considered healthy individuals with no exposure to SARS-CoV-2 virus as revealed by serology tests for antibodies against SARS-CoV-2and RT-PCR. These are the true controls since they neither have any active infection nor had been infected by the virus, previously. We considered a sample to be positive for SARS-CoV-2 if overlapping transitions were present with an intensity of greater than 1.5×10^3 and a signal to noise ratio of greater than 10 for either of the two selected peptides. This value was chosen since there will be residual peaks from carry-over of the previous sample and it was observed that the median values of the peaks in blank due to carry over were 2×10^2–4.5×10^2 depending on the transition with a maximum intensity. Thus, to negate the carry-over effect we only considered peaks with intensity more than three times of this (1.5×10^3). Using this criterion; we could distinguish SARS-CoV-2 virus-infected individuals from uninfected ones with high specificity (100%) and sensitivity (90.4%) when compared with the RT-PCR. Out of the 63 RT-PCR positive samples, we reported 57 to be positive. Of the six samples that we did not confirm positive, three satisfied the criteria of at least three transitions in one peptide to be overlapping and having an intensity of greater than 1.5×103. However, the signal to noise ratio for one of the three transitions was not above 10. In one sample although the signal to noise ratio in three transitions were more than 10, the intensity of one of the transition was 1.42×103 which fell marginally short of our criteria. In the remaining two samples, more than one transition did not meet our criteria. It should be noted that the results of RT PCR could also be falsely positive in these cases.

There are several advantages of using mass spectrometer for the diagnosis of SARS-CoV-2. The swab samples can be directly collected in vials containing inactivation buffer ensuring safe transport and handling of the samples. The major advantage of our method is its short run time with a 2.3 min gradient which is the fastest reported till date for detection of SARS-CoV-2 peptides. Cardozo et al. (2020) reported a MRM with a gradient of 10.5 min that could detect SARS-CoV-2 peptides with a sensitivity of 83%. Interestingly, they used fully automated sample preparation protocol using robotic liquid handler and turbulent flow chromatography, multiplexed for online sample clean-up and UPLC separation that enabled them to run 4 samples within the 10.5 min run. However, this kind of setup will only be available in state-of-the-art laboratories and hence cannot be used for regular diagnostic purpose. In contrast, our method can be used in any laboratory equipped with a triple quadrupole mass spectrometer.

One of the perceived disadvantages of using mass spectrometry-based diagnostics is the cost of the equipment, however, the per-sample assay cost (less than $3 per sample) is considerably lower when compared to other methods like RT-PCR. We thus, propose that targeted mass spectrometry should currently be explored for screening and diagnostic purpose, which can augment the number of tests that are currently being carried out to detect SARS-CoV-2 and complement traditional RT-PCR based assays or act as an alternate and accurate diagnostic tool with high specificity and sensitivity.

References

Bruning AHL, Leeflang MMG, Vos J et al (2017) Rapid tests for influenza, respiratory syncytial virus, and other respiratory viruses: a systematic review and meta-analysis. Clin Infect Dis 65:1026–1032

Cardozo KHM, Lebkuchen A, Okai GG et al (2020) Fast and low-cost detection of SARS-CoV-2 peptides by tandem mass spectrometry in clinical samples. https://doi.org/10.21203/rs.3.rs-28883/v1.

Corman VM, Landt O, Kaiser M et al (2020) Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 25:1

Foster MW, Gerhardt G, Robitaille L et al (2015) Targeted proteomics of human metapneumovirus in clinical samples and viral cultures. Anal Chem 87:10247–10254

Gouveia D, Grenga L, Gaillard JC, Gallais F et al (2020a) Shortlisting SARS-CoV-2 peptides for targeted studies from experimental data-dependent acquisition tandem mass spectrometry data. Proteomics 20:e2000107. https://doi.org/10.1002/pmic.202000107

Gouveia D, Miotello G, Gallais F, Gaillard J et al (2020b) Proteotyping SARS-CoV-2 virus from nasopharyngeal swabs: a proof-of-concept focused on a 3 min mass spectrometry window. J Proteome Res. https://doi.org/10.1101/2020.06.19.161000

https://www.worldometers.info/coronavirus. Accessed 18 July 2020

Iihling C, Tänzler D, Hagemann S et al (2020) Mass spectrometric identification of SARS- CoV-2 proteins from gargle solution samples of COVID-19 patients. J Proteome Res. https://doi.org/10.1021/acs.jproteome.0c00280

Majchrzykiewicz-Koehorst JA, Heikens E, Trip H et al (2015) Rapid and generic identification of influenza A and other respiratory viruses with mass spectrometry. J Virol Methods 213:75–83

Pastorino B, Touret F, Gilles M, Lamballerie XD, Charrel RN (2020) Evaluation of heating and chemical protocols for inactivating SARS-CoV-2. https://doi.org/10.1101/2020.04.11.036855

Santana WI, Williams TL, Winne EK et al (2014) Quantification of viral proteins of the avian H7 subtype of influenza virus: an isotope dilution mass spectrometry method applicable for producing more rapid vaccines in the case of an influenza pandemic. Anal Chem 86:4088–4095

Tahamtan A, Ardebili A (2020) Real-time RT-PCR in COVID-19 detection: issues affecting the results. Expert Rev Mol Diagn 20:453–454

Wang C, Horby PW, Hayden FG et al (2020) A novel coronavirus outbreak of global health concern. Lancet 395:470–473

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191. https://doi.org/10.1093/bioinformatics/btp03313