

# Targeted Detection of SARS-CoV-2 Nucleocapsid Sequence Variants by Mass Spectrometric Analysis of Tryptic Peptides

Anthony Maus, Santosh Renuse, Jennifer Kemp, Anil K. Madugundu, Patrick M. Vanderboom, Joseph Blommel, Calvin Jerde, Surendra Dasari, Benjamin R. Kipp, Ravinder J. Singh, Stefan K. Grebe, and Akhilesh Pandey\*



Cite This: *J. Proteome Res.* 2022, 21, 142–150



Read Online

ACCESS |



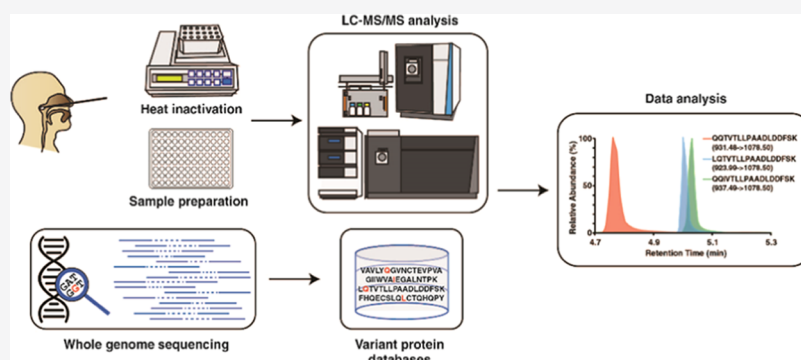
Metrics & More



Article Recommendations



Supporting Information



**ABSTRACT:** COVID-19 vaccines are becoming more widely available, but accurate and rapid testing remains a crucial tool for slowing the spread of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) virus. Although the quantitative reverse transcription-polymerase chain reaction (qRT-PCR) remains the most prevalent testing methodology, numerous tests have been developed that are predicated on detection of the SARS-CoV-2 nucleocapsid protein, including liquid chromatography-tandem mass spectrometry (LC-MS/MS) and immunoassay-based approaches. The continuing emergence of SARS-CoV-2 variants has complicated these approaches, as both qRT-PCR and antigen detection methods can be prone to missing viral variants. In this study, we describe several COVID-19 cases where we were unable to detect the expected peptide targets from clinical nasopharyngeal swabs. Whole genome sequencing revealed that single nucleotide polymorphisms in the gene encoding the viral nucleocapsid protein led to sequence variants that were not monitored in the targeted assay. Minor modifications to the LC-MS/MS method ensured detection of the variants of the target peptide. Additional nucleocapsid variants could be detected by performing the bottom-up proteomic analysis of whole viral genome-sequenced samples. This study demonstrates the importance of considering variants of SARS-CoV-2 in the assay design and highlights the flexibility of mass spectrometry-based approaches to detect variants as they evolve.

**KEYWORDS:** nucleocapsid, SARS-CoV-2, single amino acid protein variants, tandem mass spectrometry

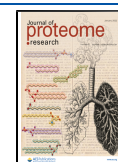
## INTRODUCTION

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) virus responsible for COVID-19 disease has infected more than 191 million people globally and resulted in more than 4.1 million deaths as of July 2021.<sup>1</sup> Access to accurate and rapid diagnostic testing is vital for case identification and public health management. Molecular testing targeting viral nucleic acids has been the primary tool for the diagnosis of COVID-19 cases.<sup>2</sup> Immunoassay-based antigen tests, such as lateral flow assays, have also emerged as a viable methodology with several tests currently authorized for use by the FDA.<sup>3</sup> However, these tests require specialized reagents, which can be difficult to obtain and the performance of these tests is highly variable.<sup>4,5</sup>

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) is already employed for the routine clinical measurement of a variety of biomolecules such as drugs, metabolites, and proteins. The pandemic has prompted many groups to explore the potential of LC-MS/MS for detection of SARS-CoV-2 as an alternative antigen testing methodology.<sup>6–13</sup> The available data support the potential of LC-

Received: July 23, 2021

Published: November 15, 2021



MS/MS to achieve >90% sensitivity and nearly 100% specificity when using quantitative reverse transcription-polymerase chain reaction (qRT-PCR) testing as the benchmark.<sup>6,10,13</sup>

As SARS-CoV-2 mutations have emerged, including those that are more contagious or partially evade host immunity,<sup>14–16</sup> there have been reports of gene variants that elude qRT-PCR and antigen detection assays. Thus, an additional potential clinical application of mass spectrometry for detection of protein/peptide sequence variants has become a possibility, although it has not been described yet.<sup>17–20</sup> Analogous to molecular assays and antigen immunoassay assays, LC-MS/MS assays to detect viral proteins also have the potential to result in false-negatives, as the amino acid substitutions can cause a change in mass that will result in an inability to detect the protein unless the method is designed appropriately. Therefore, it is imperative to have positive/negative decisions rely on multiple peptides or develop a methodology to detect peptide variants.

Our team has previously developed and validated a LC-MS/MS test with the potential to complement and supplement established RT-PCR and antigen immunoassay testing. We were able to achieve 98% sensitivity and 100% specificity for detecting SARS-CoV-2 based on monitoring of two viral nucleocapsid protein-derived peptides compared to RT-qPCR when testing was performed on samples that were a close representation of the true spectrum of viral loads seen in the actual patient population.<sup>13</sup> Interestingly, as we expanded our analysis to a larger number of cases using this approach, in a subset of samples, we observed discordant results from LC-MS/MS analysis of the two peptides from viral nucleocapsid protein that were targeted in the assay. Upon further investigation of these cases, we observed that one of the target peptides was detected at a very high abundance but the other peptide was not detected at all. To resolve this, we undertook viral genome sequencing and additional proteomic analyses and found that these discrepancies arose due to single nucleotide polymorphisms resulting in single amino acid substitutions, which altered the sequences of the targeted peptides. A minor modification to the instrument methods permitted these variants to be detected, demonstrating the power of carefully designed targeted LC-MS/MS assays to detect viral variants.

## ■ EXPERIMENTAL METHODS

### Chemicals and Reagents

We obtained phosphate-buffered saline (PBS) from Bio-Rad (Hercules, CA). Trifluoroacetic acid (TFA) from ThermoFisher Scientific (Waltham, MA), Zwittergent Z3–16 from CalBiochem (EMD Millipore, Billerica, MA), isotopically labeled AYNVTQAFGR and QQTVTLLPAADLDDFSK peptides as internal standards (IS) from New England Peptide (Gardner, MA), and an antinucleocapsid monoclonal antibody from Sino Biological (Wayne, PA, Cat# 40143-R001), which was coupled to custom MSIA D.A.R.T.'S (ThermoFisher).

### Preparation of Nasopharyngeal Swab Samples and In-Solution Trypsin Digestion

All clinical samples were deidentified prior to analysis. Nasopharyngeal (NP) swab samples were collected in PBS and 750  $\mu$ L of the sample was transferred to a 96-well plate. The virus was inactivated by adding 15  $\mu$ L of Z3–16 and

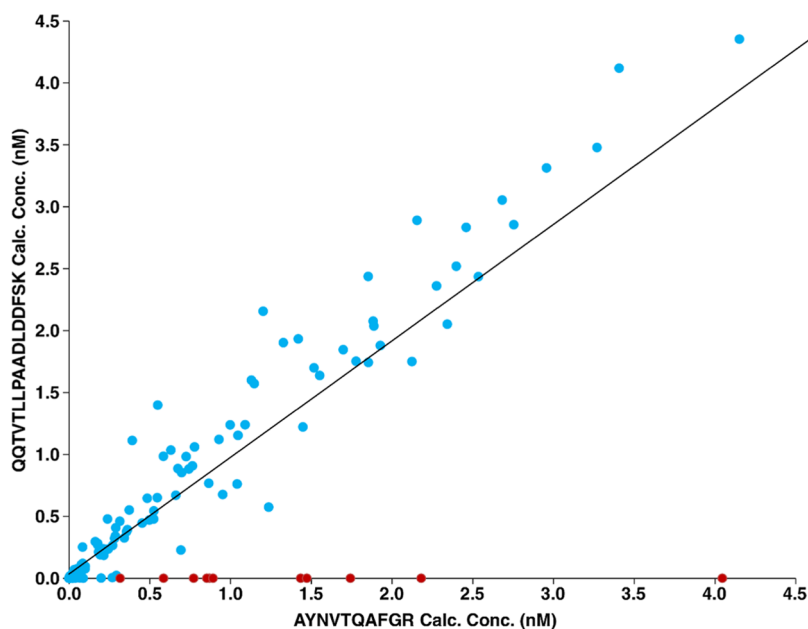
incubating at 70 °C for 30 min. Following a 10 min cooling period at 4 °C, antibody-based purification was performed using the antinucleocapsid protein monoclonal antibody coupled to MSIA D.A.R.T.'S. The purification procedure was conducted using an automated Versette liquid handling system. The tips were first washed with 1X PBS, and then the nucleocapsid protein was captured over a period of 1.75 h. Following capture, the tips were washed twice with 300  $\mu$ L of 1X PBS and then with 300  $\mu$ L of water. The nucleocapsid protein was eluted with 100  $\mu$ L of 0.2% TFA and 0.002% Z3–16 in water. The purified sample was immediately trypsin digested (rapid digest kit, Catalog#VA1060, Promega, Madison, WI). The sample eluent was mixed with 300  $\mu$ L of digest buffer followed by the addition of 1  $\mu$ g of trypsin and incubation at 70 °C for 1 h. The digestion was stopped by adding TFA to a final concentration of 1% and isotopically labeled internal standards were added. The digests were then loaded onto Evtips (EvoSep Inc., Odense, Denmark) as per manufacturer's instructions. Briefly, the C<sub>18</sub> Evtips were activated using 20  $\mu$ L of 0.1% formic acid in 100% acetonitrile followed by equilibration with 20  $\mu$ L of 0.1% formic acid in water. Activation and equilibration were carried out at 700g for 1 min using a Benchtop centrifuge. The sample was loaded at 500g for 5 min followed by washing using 0.1% formic acid once. Finally, the tips were loaded with 100  $\mu$ L of 0.1% formic acid and processed for targeted mass spectrometry analysis.

### Targeted Parallel Reaction Monitoring (PRM) Analysis

Parallel reaction monitoring (PRM) analysis was performed on an Exploris 480 mass spectrometer (ThermoFisher Scientific, San Jose, CA) interfaced with a preformed gradient LC system (EvoSep One, EvoSep Inc.). Peptides were eluted at a flow rate of 2  $\mu$ L/min and peptide separation was carried out using a 4 cm analytical column (Dr. Maisch C<sub>18</sub>AQ, 1.9  $\mu$ m, 150  $\mu$ m  $\times$  4 cm) with a 5.6 min gradient. Data acquisition parameters included MS1 scan from  $m/z$  560–1000 at a resolution of 60 000 followed by retention time scheduled PRM analysis of AYNVTQAFGR and QQTVTLLPAADLDDFSK peptides and corresponding IS peptides. The PRM parameters included an Orbitrap resolution of 60 000, an AGC target value of  $5 \times 10^4$ , an injection time of 118 ms, an isolation window of  $m/z$  1, and higher-energy collisional dissociation (HCD) with a normalized collision energy of 27. After initial screening for peptide variants, the precursor  $m/z$  corresponding to the variants LQTVTLLPAADLDDFSK ( $m/z$  923.9935) and QQIVTLLPAADLDDFSK ( $m/z$  937.4989) was added to the PRM analysis.

### Calibration and Quality Control

The recombinant SARS-CoV-2 nucleocapsid protein (97–077) purchased from ProSci (Fort Collins, CO) was used to make calibrators. Calibrators were made by spiking the recombinant protein into pooled qRT-PCR negative NP swabs in PBS at concentrations of 1, 2.5, 5, 25, and 100 pM. Given the sensitivity challenges of this type of testing, the calibration curve was designed to focus on low-concentration samples instead of attempting to interpolate concentrations of the entire patient population. Quality control samples were prepared by pooling qRT-PCR negative NP swabs (negative QC), qRT-PCR samples with cycle thresholds (Cts) of 31 (low QC), 27 (medium QC), and 25 (high QC). Calibrators



**Figure 1.** Comparison of the calculated concentrations for the AYNVTQAFGR and QQTVTLLPAADLDDFSK target peptides. The calculated concentrations of AYNVTQAFGR and QQTVTLLPAADLDDFSK target peptides are plotted as indicated (blue). The samples with no detectable QQTVTLLPAADLDDFSK signal are indicated in red.

and QC were purified, digested, and analyzed as described above.

### Mass Spectrometry Data Processing

The PRM spectra were used for subsequent data analysis. First, data were imported into Skyline<sup>21</sup> and fragment ion chromatograms were manually integrated. Next, 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 to maximize the detection 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 as described previously.<sup>22</sup> 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 distributions. This machine learning method was used to determine whether samples were positive or negative by mass spectrometry testing. Calibration curves were generated to determine protein concentrations for each target peptide in the samples analyzed. Quantitation was performed by summing the  $b_2$ ,  $y_4$ ,  $y_5$ ,  $y_6$ ,  $y_7$ , and  $y_8$  fragments of AYNVTQAFGR and the  $y_5$ ,  $y_6$ ,  $y_7$ ,  $y_8$ ,  $y_{10}$ ,  $y_{11}$ ,  $y_{12}$ , and  $y_{13}$  fragments of QQTVTLLPAADLDDFSK (and variants). A custom R script was used to correct the summed intensity of the analyte fragments with the equivalent fragments of the isotopically labeled internal standards and generate calibration curves for AYNVTQAFGR and QQTVTLLPAADLDDFSK from the calibrators using  $1/x$  weighting. The resulting linear regression equation was used to determine protein concentrations from each peptide in QC and patient samples. Additional analysis and plot generation were performed using Microsoft Excel.

### Processing of Nasopharyngeal Swab Samples for Untargeted LC-MS/MS Analysis

Nasopharyngeal swab samples were collected in PBS and processed for in-solution trypsin digestion. One milliliter of the sample was reduced using 10 mM TCEP at RT for 20 min followed by alkylation using 20 mM IAA at RT in dark for 30 min. 5  $\mu$ g of sequencing grade trypsin + Lys-C mix was added to the samples and incubated overnight at 37 °C. The samples were desalted using  $C_{18}$  solid-phase extraction spin columns (Glygen, Columbia, MD), dried down, and reconstituted in 20  $\mu$ L of 0.2% formic acid. Samples were split into three parts—10% for single-shot discovery, 10% for single-shot PRM, and the remaining 80% was fractionated into 6 SCX fractions for DDA analysis. SCX fractionation was carried out using the protocol as described previously.<sup>23</sup> Briefly, the samples were reconstituted in 1% TFA and loaded on SCX top tips containing PolySULFOETHYL A using the Benchtop centrifuge at 1000 rpm for 1 min. Loaded peptides were washed using 0.2% TFA twice and sequentially eluted into 6 fractions using 50, 75, 100, 150, 200, 300 mM Ammonium acetate and 5%  $NH_4OH$  in 80% acetonitrile. The eluted fractions were vacuum dried till further analysis.

LC-MS/MS analysis for targeted and untargeted discovery proteomics experiments was carried out using an Ultimate 3000 RSLCnano system (ThermoFisher Scientific) connected to an Orbitrap Exploris 480 mass spectrometer. Single-shot PRM and DDA analyses were carried out for detection of variant peptides from SARS-CoV-2. Sample-specific mutant peptide targeted lists were created including the maximum of three missed cleavages and were analyzed on the Exploris 480 mass spectrometer using the PRM method. Samples were also analyzed using an untargeted single-shot DDA method. 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) with a flow rate of 300 nL/min and a linear

**Table 1. Genome Sequence and Targeted LC-MS/MS Results for Samples with High AYNVTQAFGR Concentrations and No Detectable QQTVTLLPAADLDDFSK**

sample number	C <sub>p</sub> (qRT-PCR)	amino acid substitution (based on genome sequencing)	AYNVTQAFGR concentration (pM)	QQTVTLLPAADLDDFSK concentration (pM)	LQTVTLLPAADLDDFSK concentration (pM)	QQIVTLLPAADLDDFSK concentration (pM)
29	27.66	p.Gln389Leu	3775	<LLOQ	1797.653487	<LLOQ
48	28.59	p.Gln389Leu	507	<LLOQ	322.8341255	<LLOQ
205	28.33	Unable to Sequence	988	<LLOQ	<LLOQ	849.2092848
217	23.42	p.Gln389Leu	2046	<LLOQ	1201.033373	<LLOQ
221	24.51	p.Thr391Ile	959	<LLOQ	<LLOQ	705.7093037
235	22.76	p.Gln389Leu	1974	<LLOQ	740.6691727	<LLOQ
237	23.49	p.Gln389Leu	1041	<LLOQ	939.0091282	<LLOQ
255	24.89	p.Thr391Ile	798	<LLOQ	<LLOQ	879.7508931
257	25.09	p.Thr391Ile	187	<LLOQ	<LLOQ	182.7447412
281	25.31	p.Thr391Ile	2518	<LLOQ	<LLOQ	3295.20967
312	26.05	p.Thr391Ile	4266	<LLOQ	<LLOQ	3374.362533

gradient of 5 to 40% solvent B (100% ACN, 0.1% formic acid) over 70 min. Both precursor and fragment ions were acquired using an Orbitrap mass analyzer. Precursor ions were acquired in an  $m/z$  range of 350–1800 with a resolution of 120 000 (at  $m/z$  200). Precursor fragmentation was carried out using the HCD method with a normalized collision energy (NCE) of 27. The fragment ions were acquired at a resolution of 30,000 (at  $m/z$  200). The scans were arranged using a top-speed method with a cycle time of 3 s between MS and MS/MS. Ion transfer capillary voltage was maintained at 1.9 kV. For internal mass calibration, the lock mass option was enabled with polysiloxane ions ( $m/z$  445.120025) from ambient air.

#### Mass Spectrometry Data Analysis of Untargeted LC-MS/MS Data

The raw mass spectrometry data were searched using Andromeda<sup>24</sup> in the MaxQuant software suite (version 1.6.7.0) against a combined protein database containing the UniProt human protein database and observed mutant SARS-CoV-2 protein sequences based on the genome sequencing data for each sample including common MS contaminants. The raw data were searched against the sample-specific variant SARS-CoV-2 protein database. The search parameters included a maximum of two missed cleavages: carbamidomethylation at cysteine as a fixed modification for samples that were reduced and alkylated, and 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. The false discovery rate was set to 1% at peptide-spectrum matches (PSMs) and peptide and protein levels.

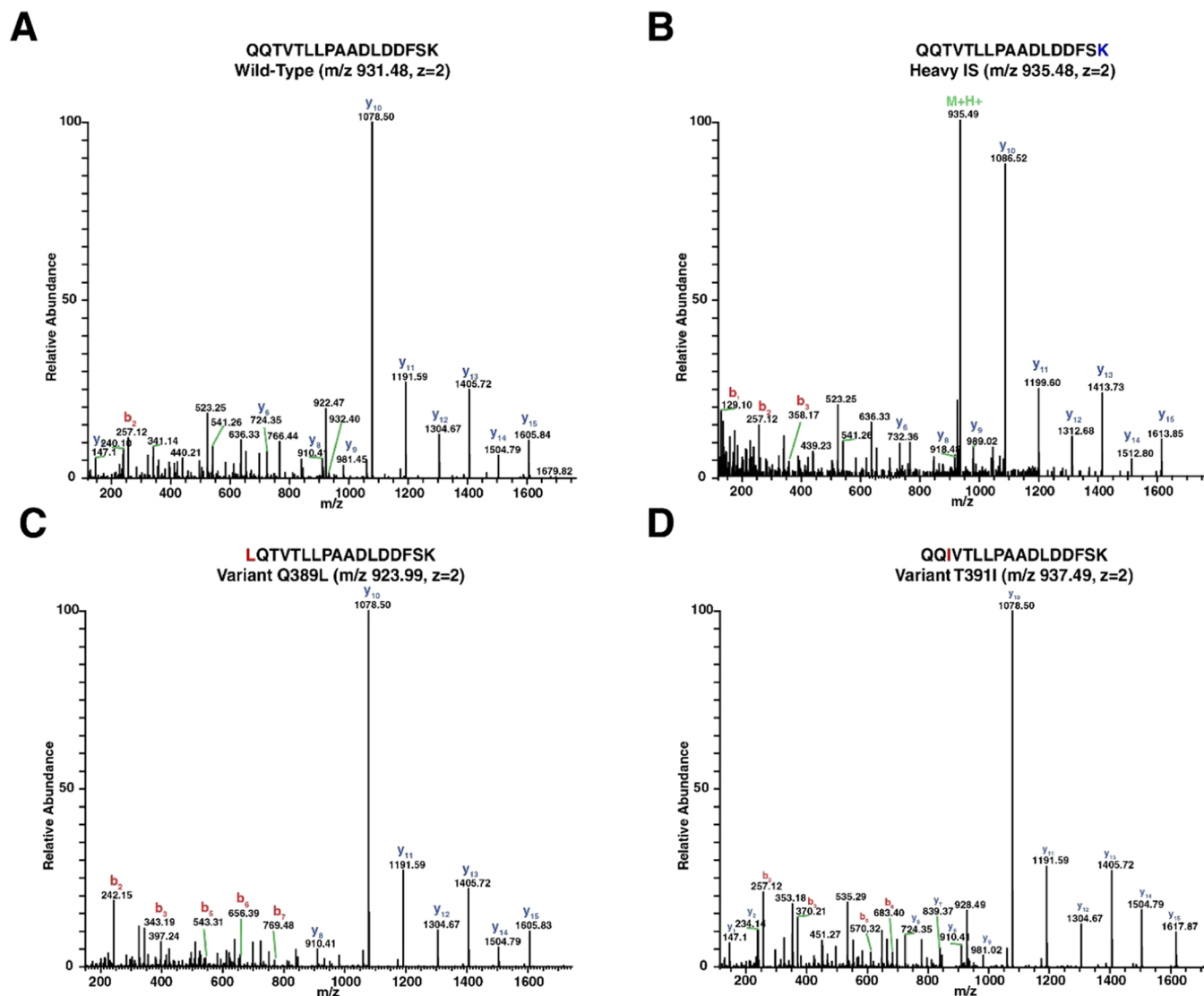
#### Genome Sequencing of SARS-CoV-2 Samples

SARS-CoV-2 extraction was performed using the Mag-MAXTM viral/pathogen nucleic acid isolation kit (Cat# A42352) from Applied Biosystems (ThermoFisher Scientific). Viral RNA extraction was performed following manufacturer's protocol, using 400  $\mu$ L of sample transport media. RNA was eluted in 55  $\mu$ L of which 25  $\mu$ L was loaded onto the sequencing instrument. Sequencing was performed using the Ion AmpliSeq SARS-CoV-2 Research Panel for Genexus 6.2 from ThermoFisher Scientific following manufacturer's instructions.

## RESULTS AND DISCUSSION

We recently completed a study to evaluate multiple testing platforms for the analysis of 350 clinical nasopharyngeal swab samples including LC-MS/MS, point-of-care tests, qRT-PCR, and digital droplet PCR.<sup>4</sup> The samples included 250 SARS-CoV-2 positive samples and 100 negative samples as determined by qRT-PCR. Figure 1 shows a comparison of the concentration results for the AYNVTQAFGR and QQTVTLLPAADLDDFSK target peptides. The correlation is imperfect in part because the concentration of many of these samples had concentrations above the calibration curve (top calibrator = 100 pM); but the majority of samples produced the expected results of equimolar concentrations for the two target peptides. However, we observed 11 samples with a high concentration of the AYNVTQAFGR peptide, while the abundance of QQTVTLLPAADLDDFSK was below the lower limit of quantitation (LLOQ). To investigate if the peptide was not identified because of a sequence variation, viral genome sequencing was performed for these 11 samples. The FASTA files of the genome sequences are shown in the Supporting Information. As listed in Table 1, single nucleotide polymorphisms (SNPs) causing corresponding single amino acid substitutions that alter the sequence of the QQTVTLLPAADLDDFSK target peptide were confirmed by sequencing in 10/11 samples. One sample was unable to be sequenced due to fewer viable RNA copies, which can be attributed to lower viral load or degradation through the sample handling processes and is evident by the lower sequencing read metrics, as shown in Figure S1. Figure S2 shows multiple sequence alignment of nucleocapsid protein sequences derived from the 10 samples that were successfully sequenced indicating the position of variants. We observed two variants in the peptide—QQTVTLLPAADLDDFSK-Q389L (LQTVTLLPAADLDDFSK) and T391I (QQIVTLLPAADLDDFSK), both of which resulted in an alteration of precursor ion mass, which explained our inability to detect the second target peptide.

The PRM technique used to maximize the sensitivity of our test targeted only the precursor ion masses of the wild-type peptide sequences ( $m/z$  563.7856 and 931.4807), while a modified PRM method enabled detection of the LQTVTLLPAADLDDFSK and QQIVTLLPAADLDDFSK variants by adding their precursor ion masses ( $m/z$  923.9935 and 937.4989, respectively) to the target list. Annotated MS/MS spectra are shown in Figure 2 for the wild type (Figure 2A),

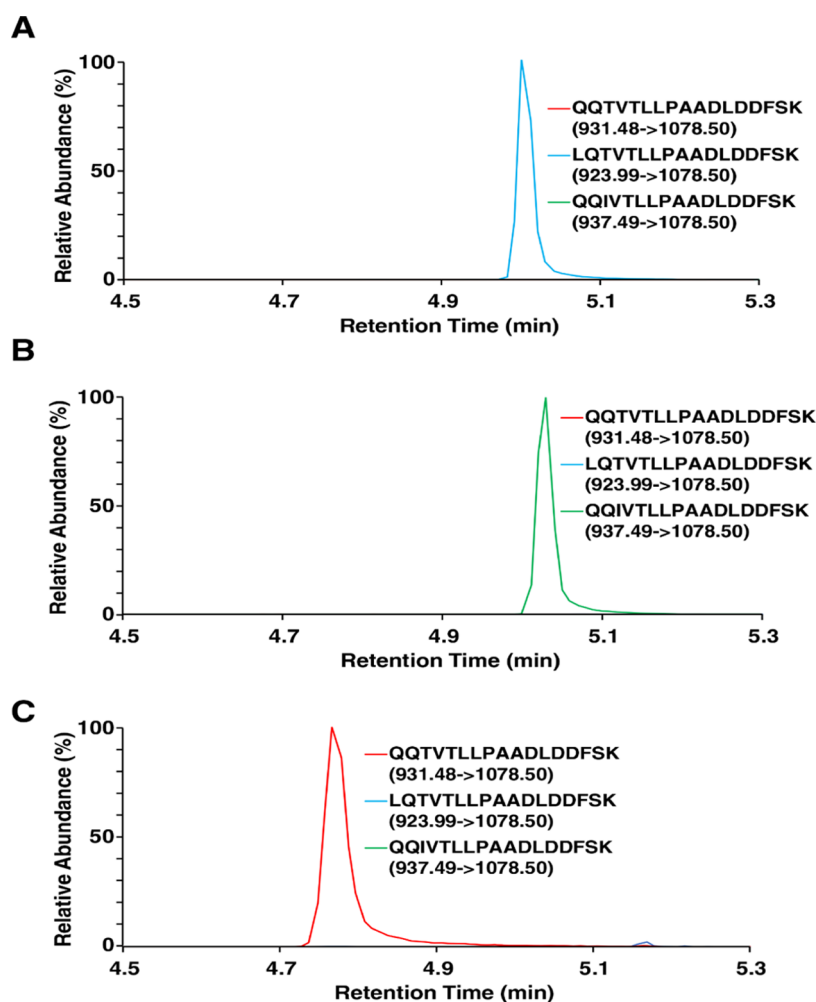


**Figure 2.** Annotated MS/MS spectra of: (A) wild-type QQTVTLLPAADLDDFSK, (B) isotopically labeled internal standard, (C) LQTVTLLPAADLDDFSK variant, and (D) QQIVTLLPAADLDDFSK variant.

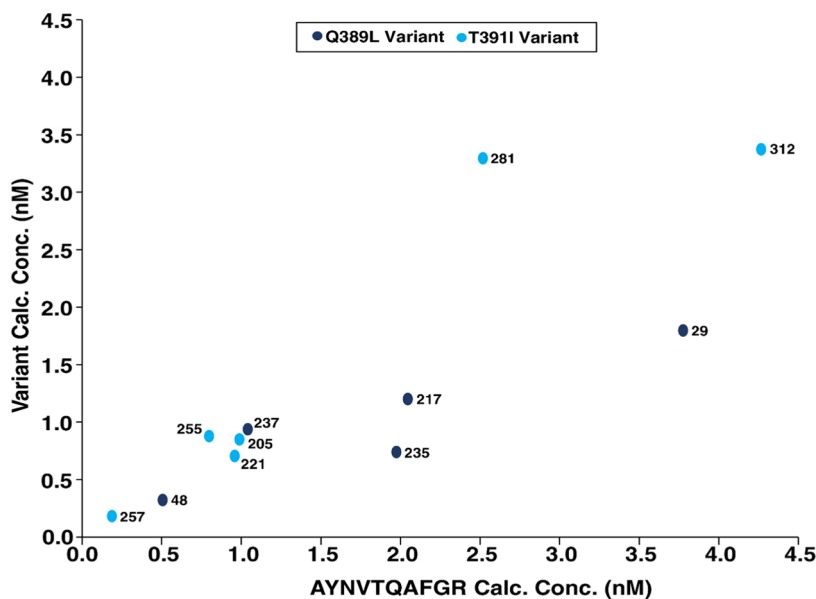
the isotopically labeled internal standard (Figure 2B), the LQTVTLLPAADLDDFSK variant (Figure 2C), and the QQIVTLLPAADLDDFSK variant (Figure 2D). Figure 3 shows extracted ion chromatograms (XICs) of the most abundant product ions for the wild-type peptide sequence and the two variants of interest. Figure 3A is a typical sample with no variants detected and a large signal for the QQTVTLLPAADLDDFSK peptide; however, as shown in Figure 3B,C, the product ions expected from the wild-type sequence were not observed, but instead the predicted variants were detected.

The concentrations of the peptides were estimated using the QQTVTLLPAADLDDFSK internal standard and the calibration curve. These results are listed in Table 1, and the linear regression plot is shown in Figure 4. Ultimately, all the predicted variants were detected with only a slight modification to the MS data acquisition method. Additionally, we were able to detect a variant in the sample where we were unable to obtain the genomic sequence, demonstrating the sensitivity and ability of this technique to be extrapolated to samples in the absence of genome sequencing once variant targets have been established.

Finally, we sought to assess the utility of LC-MS/MS for detection of other variants predicted by the genome sequences. To accomplish this, we carried out targeted and untargeted LC-MS/MS analysis, as depicted in Figure S3. Targeted PRM analyses resulted in identification of LQTVTLLPAADLDDFSK and QQIVTLLPAADLDDFSK variants in respective samples, as shown in Table 2. We detected the DGIWVAIEGALNTPK variant in samples 257 (by PRM and untargeted DDA) and 281 (PRM), while the NSTLGSRR variant could be detected in samples 48 (PRM) and 235 (SCX fractionated sample). The identification of variants of the DGIWVATEGALNTPK peptide is especially significant, as we and others have found it to be a sensitive target peptide for the LC-MS/MS analysis of SARS-CoV-2.<sup>5,11</sup> In addition to the nucleocapsid variants, we also detected the variant peptide FISTCACEIVGGQIITCAK from the protein ORF1ab in five samples. Although the peptide identifications from database searching in this study were obtained using variants identified based on genomic sequence information, it is possible that variants could also be identified in the absence of genome sequencing through error-tolerant database searching.<sup>25</sup>



**Figure 3.** Extracted ion chromatograms of the most abundant fragment ions produced by the QQTVTLLPAADLDDFSK, LQTVTLLPAADLDDFSK, and QQIVTLLPAADLDDFSK peptides. (A) Typical patient sample with only wild-type signals being detected, (B) patient sample with the LQTVTLLPAADLDDFSK variant, and (C) patient sample with the QQIVTLLPAADLDDFSK variant.



**Figure 4.** Comparison of respective variant concentrations versus AYNVTQAFGR calculated concentration.

The nucleocapsid protein is the most abundantly expressed protein by SARS-CoV-2,<sup>7-9,11,13</sup> making it an appealing target

for testing methodologies predicated on detecting viral proteins such as LC-MS/MS proteomic approaches and

**Table 2. Identification of SARS-CoV-2 Variants by Direct Digest Followed by NanoLC-MS/MS Analysis and Variant-Specific Protein Database Searches**

protein	site	peptide sequence	sample number	PRM–Skyline	identified in single-shot DDA	identified in SCX (6 fractions)		
nucleocapsid	135	DGIWVAIEGALNTPK	257	✓	✓	×		
			281	✓	×	×		
	389	LQTVTLLPAADLDDFSK	029	✓	×	×		
			048	✓	×	×		
			217	✓	×	×		
			235	✓	×	×		
			237	✓	×	×		
			235	×	×	✓		
	199	NSTLGSSR	048	✓	×	×		
	391	QQIVTLLPAADLDDFSK	221	✓	×	✓		
			255	✓	×	×		
			257	✓	×	×		
			281	✓	×	×		
			312	✓	×	×		
ORF1ab			665	FISTCACEIVGGQITCAK	221	✓	×	×
					255	✓	✓	×
	257	✓			×	×		
	281	✓			×	×		
	312	✓			✓	✓		

point-of-care tests using lateral flow immunoassay and fluorescence immunoassay techniques. Therefore, the nucleocapsid protein is of great significance for diagnostic purposes and an understanding of the prevalence of nucleocapsid variants and their implications for each methodology is imperative. When utilizing LC-MS/MS to measure a select set of target peptides, strategies to ensure accurate patient results in the presence of variants must be utilized. We have minimized the impact of variants on test results by adding the precursor ion masses of the common variants to our targeted PRM method, measuring signals from multiple peptides, and deploying a weighted machine learning approach for determining positive/negative results that are not reliant on detection of both peptides. An alternative strategy to minimize the impact of protein variants is to use a bioinformatics approach to eliminate peptides that are known to have a high prevalence of variants.<sup>4</sup> However, in doing this, one may disregard sensitive and specific biomarkers, as in the case of the QQTVTLLPAADLDDFSK peptide, which is an excellent proteotypic peptide for detection.

We must note that our methodologies are not currently optimized for detection of the wild-type spike protein nor variants of the spike protein, which have been of greater public interest due to the potential of these variants to alter viral transmission rates, disease severity, and reduced effectiveness of natural or vaccine-induced immunity.<sup>26</sup> Based on these results, it is likely that MS-based approaches could also be utilized for detection and identification of spike protein variants.

## CONCLUSIONS

The nucleocapsid protein of SARS-CoV-2 has emerged as an important biomarker for diagnostic tests based on protein detection (i.e., LC-MS/MS and immunoassays). We suspected that protein variants were responsible for our failure to detect a target peptide derived from the nucleocapsid protein in a subset of nasopharyngeal swap samples. The presence of variants was confirmed by genomic sequencing, and with only minor modifications to the LC-

MS/MS instrument method, we were able to detect the sequence variants of our target peptide, and additional variants were detected in bottom-up proteomic analyses. This work highlights the importance of developing a strategy for ensuring accurate test results in the presence of variants when utilizing protein-based approaches, especially when measuring a select set of target peptides.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.1c00613>.

Sequencing read metrics from a successfully sequenced sample and a sample that was unable to be sequenced (Figure S1); multiple sequence alignment of nucleocapsid protein sequences derived from the 10 samples that were successfully genome sequenced indicating the position of variants (Figure S2); and diagram of the workflow utilized for identification of variant peptides by LC-MS/MS (Figure S3) (PDF)

## AUTHOR INFORMATION

### Corresponding Author

**Akhilesh Pandey** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States; Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota 55905, United States; [orcid.org/0000-0001-9943-6127](https://orcid.org/0000-0001-9943-6127); Email: [Pandey.Akhilesh@mayo.edu](mailto:Pandey.Akhilesh@mayo.edu)

### Authors

**Anthony Maus** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States; [orcid.org/0000-0003-3006-6265](https://orcid.org/0000-0003-3006-6265)

**Santosh Renuse** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and

Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States; Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota 55905, United States

**Jennifer Kemp** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States

**Anil K. Madugundu** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States; Institute of Bioinformatics, Bangalore 560066 Karnataka, India; Manipal Academy of Higher Education, Manipal 576104 Karnataka, India; Center for Molecular Medicine, National Institute of Mental Health and Neurosciences, Bangalore 560029 Karnataka, India

**Patrick M. Vanderboom** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States

**Joseph Blommel** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States

**Calvin Jerde** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States

**Surendra Dasari** – Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota 55905, United States

**Benjamin R. Kipp** – Department of Laboratory Medicine and Pathology, Division of Laboratory Genetics and Genomics, Mayo Clinic, Rochester, Minnesota 55905, United States

**Ravinder J. Singh** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States

**Stefan K. Grebe** – Department of Laboratory Medicine and Pathology, Division of Clinical Biochemistry and Immunology, Mayo Clinic, Rochester, Minnesota 55905, United States; Department of Medicine, Division of Endocrinology, Mayo Clinic, Rochester, Minnesota 55902, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jproteome.1c00613>

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest. The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PMID: 27924013) with the data set identifier PXD026795 and Panorama Public (URL: <https://panoramaweb.org/covid19variants.url>). The viral genome sequences have been submitted to sequence read archive (NCBI; BioProject: PRJNA749298).

### ACKNOWLEDGMENTS

This study was supported by DBT/Wellcome Trust India Alliance Margdarshi Fellowship grant IA/M/15/1/S02023 awarded to Akhilesh Pandey and the generosity of Eric and Wendy Schmidt.

### ABBREVIATIONS USED

SARS-CoV-2, severe acute respiratory syndrome coronavirus-2; LC, liquid chromatography; MS/MS, tandem mass spectrometry; MS, mass spectrometry; RT-PCR, real-time polymerase chain reaction; PBS, phosphate-buffered saline; TFA, trifluoroacetic acid; IS, internal standard; NP, nasopharyngeal; PRM, parallel reaction monitoring; CT, cycle threshold; CV, coefficient of variation; LLOQ, lower limit of quantitation; SNP, single nucleotide polymorphism; XIC, extracted ion chromatogram

### REFERENCES

- (1) WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int> (accessed Mar 12, 2021).
- (2) Udugama, B.; Kadhiresan, P.; Kozlowski, H. N.; Malekjahani, A.; Osborne, M.; Li, V. Y. C.; Chen, H.; Mubareka, S.; Gubbay, J. B.; Chan, W. C. W. Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano* **2020**, *14*, 3822–3835.
- (3) U.S. Food & Drug Administration *Vitro Diagnostics EUAs*; FDA, 2021.
- (4) Karon, B. S.; Donato, L.; Bridgeman, A. R.; Blommel, J. H.; Kipp, B.; Maus, A.; Renuse, S.; Kemp, J.; Madugundu, A. K.; Vanderboom, P. M.; Chavan, S.; Dasari, S.; Singh, R. J.; Grebe, S. K. G.; Pandey, A. Analytical Sensitivity and Specificity of Four Point of Care Rapid Antigen Diagnostic Tests for SARS-CoV-2 Using Real-Time Quantitative PCR, Quantitative Droplet Digital PCR, and a Mass Spectrometric Antigen Assay as Comparator Methods. *Clin. Chem.* **2021**, No. hvab138.
- (5) Wagenhäuser, I.; Knies, K.; Rauschenberger, V.; Eisenmann, M.; McDonogh, M.; Petri, N.; Andres, O.; Flemming, S.; Gawlik, M.; Papsdorf, M.; Taurines, R.; Böhm, H.; Forster, J.; Weismann, D.; Weißbrich, B.; Dölken, L.; Liese, J.; Kurzai, O.; Vogel, U.; Krone, M. Clinical Performance Evaluation of SARS-CoV-2 Rapid Antigen Testing in Point of Care Usage in Comparison to RT-QPCR. *EBioMedicine*. **2021**, *69*, No. 103455.
- (6) Singh, P.; Chakraborty, R.; Marwal, R.; Radhakrishnan, V. S.; Bhaskar, A. K.; Vashisht, H.; Dhar, M. S.; Pradhan, S.; Ranjan, G.; Imran, M.; Raj, A.; Sharma, U.; Singh, P.; Lall, H.; Dutta, M.; Garg, P.; Ray, A.; Dash, D.; Sivasubbu, S.; Gogia, H.; Madan, P.; Kabra, S.; Singh, S. K.; Agrawal, A.; Rakshit, P.; Kumar, P.; Sengupta, S. A Rapid and Sensitive Method to Detect SARS-CoV-2 Virus Using Targeted-Mass Spectrometry. *J. Proteins Proteomics* **2020**, *11*, 159–165.
- (7) Cardozo, K. H. M.; Lebkuchen, A.; Okai, G. G.; Schuch, R. A.; Viana, L. G.; Olive, A. N.; Lazari, C. dos S.; Fraga, A. M.; Granato, C. F. H.; Pintão, M. C. T.; Carvalho, V. M. Establishing a Mass Spectrometry-Based System for Rapid Detection of SARS-CoV-2 in Large Clinical Sample Cohorts. *Nat. Commun.* **2020**, *11*, No. 6201.
- (8) Saadi, J.; Oueslati, S.; Bellanger, L.; Gallais, F.; Dortet, L.; Roque-Afonso, A.-M.; Junot, C.; Naas, T.; Fenaille, F.; Becher, F. 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*, 1434–1443.
- (9) Ihling, C.; Tänzler, D.; Hagemann, S.; Kehlen, A.; Hüttelmaier, S.; Arlt, C.; Sinz, A. Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. *J. Proteome Res.* **2020**, *19*, 4389–4392.
- (10) Schuster, O.; Zvi, A.; Rosen, O.; Achdout, H.; Ben-Shmuel, A.; Shifman, O.; Yitzhaki, S.; Laskar, O.; Feldberg, L. Specific and



Rapid SARS-CoV-2 Identification Based on LC-MS/MS Analysis. *ACS Omega* **2021**, *6*, 3525–3534.

(11) Gouveia, D.; Miotello, G.; Gallais, F.; Gaillard, J.-C.; Debroas, S.; Bellanger, L.; Lavigne, J.-P.; Sotto, A.; Grenga, L.; Pible, O.; Armengaud, J. 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*, 4407–4416.

(12) Cazares, L. H.; Chaerkady, R.; Samuel Weng, S. H.; Boo, C. C.; Cimbrow, R.; Hsu, H.-E.; Rajan, S.; Dall'Acqua, W.; Clarke, L.; Ren, K.; McTamney, P.; Kallewaard-LeLay, N.; Ghaedi, M.; Ikeda, Y.; Hess, S. Development of a Parallel Reaction Monitoring Mass Spectrometry Assay for the Detection of SARS-CoV-2 Spike Glycoprotein and Nucleoprotein. *Anal. Chem.* **2020**, *92*, 13813–13821.

(13) Renuse, S.; Vanderboom, P. M.; Maus, A. D.; Kemp, J. V.; Gurtner, K. M.; Madugundu, A. K.; Chavan, S.; Peterson, J. A.; Madden, B. J.; Mangalparthi, K. K.; Mun, D.-G.; Singh, S.; Kipp, B. R.; Dasari, S.; Singh, R. J.; Grebe, S. K.; Pandey, A. A Mass Spectrometry-Based Targeted Assay for Detection of SARS-CoV-2 Antigen from Clinical Specimens. *EBioMedicine* **2021**, *69*, No. 103465.

(14) Wang, R.; Chen, J.; Gao, K.; Hozumi, Y.; Yin, C.; Wei, G.-W. Analysis of SARS-CoV-2 Mutations in the United States Suggests Presence of Four Substrains and Novel Variants. *Commun. Biol.* **2021**, *4*, 1–14.

(15) Li, Q.; Wu, J.; Nie, J.; Zhang, L.; Hao, H.; Liu, S.; Zhao, C.; Zhang, Q.; Liu, H.; Nie, L.; Qin, H.; Wang, M.; Lu, Q.; Li, X.; Sun, Q.; Liu, J.; Zhang, L.; Li, X.; Huang, W.; Wang, Y. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **2020**, *182*, 1284–1294.e9.

(16) Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J. C.; Muecksch, F.; Rutkowska, M.; Hoffmann, H.-H.; Michailidis, E.; Gaebler, C.; Agudelo, M.; Cho, A.; Wang, Z.; Gazumyan, A.; Cipolla, M.; Luchsinger, L.; Hillyer, C. D.; Caskey, M.; Robbiani, D. F.; Rice, C. M.; Nussenzweig, M. C.; Hatziioannou, T.; Bieniasz, P. D. Escape from Neutralizing Antibodies by SARS-CoV-2 Spike Protein Variants. *eLife* **2020**, *9*, No. e61312.

(17) Zhang, B.; Whiteaker, J. R.; Hoofnagle, A. N.; Baird, G. S.; Rodland, K. D.; Paulovich, A. G. Clinical Potential of Mass Spectrometry-Based Proteogenomics. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 256–268.

(18) Alfaro, J. A.; Ignatchenko, A.; Ignatchenko, V.; Sinha, A.; Boutros, P. C.; Kislinger, T. Detecting Protein Variants by Mass Spectrometry: A Comprehensive Study in Cancer Cell-Lines. *Genome Med.* **2017**, *9*, 1–12.

(19) Macklin, A.; Khan, S.; Kislinger, T. Recent Advances in Mass Spectrometry Based Clinical Proteomics: Applications to Cancer Research. *Clin. Proteomics* **2020**, *17*, 1–25.

(20) Maus, A.; Kemp, J.; Milosevic, D.; Renuse, S.; Pandey, A.; Singh, R. J.; Grebe, S. K. G. Center of Mass Calculation in Combination with MS/MS Allows Robust Identification of Single Amino Acid Polymorphisms in Clinical Measurements of Insulin-Like Growth Factor-1. *J. Proteome Res.* **2020**, *19*, 186–193.

(21) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* **2010**, *26*, 966–968.

(22) van der Laan, M. J.; Polley, E. C.; Hubbard, A. E. Super Learner. *Stat. Appl. Genet. Mol. Biol.* **2007**, *6* (1), 25.

(23) Madugundu, A. K.; Na, C. H.; Nirujogi, R. S.; Renuse, S.; Kim, K. P.; Burns, K. H.; Wilks, C.; Langmead, B.; Ellis, S. E.; Collado-Torres, L.; Halushka, M. K.; Kim, M.-S.; Pandey, A. Integrated Transcriptomic and Proteomic Analysis of Primary Human Umbilical Vein Endothelial Cells. *Proteomics* **2019**, *19*, No. 1800315.

(24) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine

Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10*, 1794–1805.

(25) Creasy, D.; Cottrell, J. Error tolerant searching of uninterpreted tandem mass spectrometry data. *Proteomics* **2002**, *2*, 1426–1434.

(26) CDC. Coronavirus Disease 2019 (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html> (accessed May 09, 2021).