Article

Identification of Endogenous Peptides in Nasal Swab Transport Media used in MALDI-TOF-MS Based COVID-19 Screening

Helen Tsai, Brett S. Phinney,* Gabriela Grigorean, Michelle R. Salemi, Hooman H. Rashidi, John Pepper, and Nam K. Tran

Read Online

ACCESS | 📊 Metrics & More | 📖 Article Recommendations | SI Supporting Information

ABSTRACT: Mass spectrometry (MS) based diagnostic detection of 2019 novel coronavirus infectious disease (COVID-19) has been postulated to be a useful alternative to classical PCR based diagnostics. These MS based approaches have the potential to be both rapid and sensitive and can be done on-site without requiring a dedicated laboratory or depending on constrained supply chains (i.e., reagents and consumables). Matrix-assisted laser desorption ionization (MALDI)−time-of-flight (TOF) MS has a long and established history of microorganism detection and systemic disease assessment. Previously, we have shown that automated machine learning (ML) enhanced MALDI-TOF-MS screening of nasal swabs can be both sensitive and specific for COVID-19

detection. The underlying molecules responsible for this detection are generally unknown nor are they required for this automated ML platform to detect COVID-19. However, the identification of these molecules is important for understanding both the mechanism of detection and potentially the biology of the underlying infection. Here, we used nanoscale liquid chromatography tandem MS to identify endogenous peptides found in nasal swab saline transport media to identify peptides in the same the mass over charge (m/z) values observed by the MALDI-TOF-MS method. With our peptidomics workflow, we demonstrate that we can identify endogenous peptides and endogenous protease cut sites. Further, we show that SARS-CoV-2 viral peptides were not readily detected and are highly unlikely to be responsible for the accuracy of MALDI based SARS-CoV-2 diagnostics. Further analysis with more samples will be needed to validate our findings, but the methodology proves to be promising.
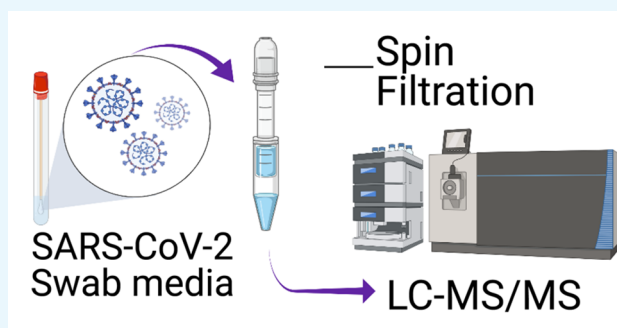
INTRODUCTION

The first known case of novel coronavirus infectious disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) was first identified in Wuhan, China, in December 2019. The disease has since quickly escalated into a global pandemic. In response, scientists from around the world have generated considerable research that has led to a better understanding of SARS-CoV-2 and the management of COVID-19. Unfortunately, we still face problems with limiting the spread of COVID-19 and highly infectious variants. Rapid, on-site, and robust screening for SARS-CoV-2 infection could enhance both containment and reduction in infectivity. A rapid on-site test and constant screening can help determine the social restriction policies, track new variants and their spread, and assess treatments. For a rapid on-site test to be viable, it will need to have adequate sensitivity and specificity, as well as yield low false-positive rate (FPR) and false-negative rate (FNR).

There are many methodologies for detecting SARS-CoV-2 including molecular and antigen approaches. Molecular methods, such as reverse transcription polymerase chain reaction (RT-PCR), are the accepted gold standard. These

molecular methods are highly sensitive and specific and can be automated to provide high throughput testing capacity. However, these molecular methods often require a specialized laboratory and reagents that can be in short supply and significant infrastructure for the transportation of samples. Results are typically produced in 24 to 48 h. Point-of-care (POC) molecular methods exist and can report results in as little as 20 min, but these platforms are not widely available and are often impacted by supply chains. Antigen methods are rapid and low-cost alternatives to molecular methods. Both POC and laboratory-based antigen tests are now available but appear to be less sensitive and specific compared to their molecular counterparts.

Recently, mass spectrometry (MS) has been used as an alternative to RT-PCR detection, using both liquid chromatog-

Table 1. Metadata of the Samples Used

sample	days since symptom onset	age	diagnosis	LIAT result	public health/UCD result	ddPCR result	ddPCRN1 (copies/μL)	ddPCRN2 (copies/μL)
n13		70	intertrochanteric fracture of left femur	negative	control	negative	0	0
n5	1	64	COPD; pulmonary emphysema	negative	control	negative	0	0
n6	4	63	COPD exacerbation	negative	control	negative	0	0
n9	2	31	moderate asthma w/acute exacerbation	negative	control	negative	0	0
p22	1	30		positive	COVID	positive	6.91	7.1
p25	10	77		positive	COVID	positive	8.78	6.25
p30	0	68		positive	COVID	positive	1.14	0.663
p32	2	37		positive	COVID	positive	0.683	0.683

raphy tandem MS (LC-MS/MS)[1−5] and matrix assisted laser desorption ionization (MALDI)−time-of-flight (TOF) MS.[6−9] MS approaches generally do not rely on reagents that can be in short supply or are biologically produced, with the exception of trypsin for bottom up proteomics methods, and have a long track record of successful microorganism identification[10−13] and effective assessment of systemic disease.[14,15] While LC-MS based approaches can be fast (1−5 min) and have high sensitivity and accuracy,[1] they typically require complex instrumentation that requires both dedicated laboratory facilities and highly trained personnel. In contrast, MALDI-TOF-MS approaches can be performed on-site and generally do not require infrastructure such as specialized laboratories and highly trained personnel.

MALDI-TOF-MS based techniques have a long-proven track record in clinical microbiology for pathogen identification. These MALDI-TOF-MS based approaches rely on "spectral patterns" of generally unknown components to diagnose disease and detect microorganisms. Recently, Tran et al. demonstrated that machine learning (ML)-enhanced MALDI-TOF-MS screening of SARS-CoV-2 nasal swabs can be both accurate and sensitive.[6] Due to the limitation of MALDI-TOF-MS technology, the underlying molecules responsible for the spectra are unknown. Identification of these components will be useful for the understanding of both the mechanism of detection and the underlying biology. We followed up with an exploratory study using nanoscale LC-MS/MS to identify the underlying peptides that could be responsible for the m/z values seen in the MALDI-TOF spectra. At the onset of the study, our exploratory investigation has the following limitations: (1) nanoscale LC-MS/MS is far more sensitive and can detect more peptides than MALDI-TOF-MS, so direct identification cannot be made but only inferred; (2) our method will only identify potential peptide components of the spectra but will miss other molecules such as lipids and carbohydrates; (3) our sample size is limited so the study serves as a template for future studies.

We hypothesize that the peptides attached to the exterior of the nasal swabs used in the MALDI-TOF-MS study are digested by endogenous proteases. This is supported by the mass range in the MALDI-TOF spectrum. Thus, we chose to perform our investigation using a peptidomics workflow, instead of a trypsin-based proteomics workflow, because this method also relies on endogenous proteases for digestion. We believe that our peptidomics workflow can be applied to the nasal swab transport media to identify host proteome profiles. The identification of the nasal endogenous peptides during infection can help us further understand SARS-CoV-2

pathogenesis and determine suitable detection methods and discover drug targets.

Here, we show that the peptidomics workflow is suitable for the identification of peptides in nasal swab saline transport media. We identified endogenous protease cut sites and 14720 endogenous peptides where the top proteins mapped are comprised of polymeric immunoglobulin receptor, actin, statherin, glyceraldehyde-3-phosphate dehydrogenase, thymosin β-4, and histones. We show that SARS-CoV-2 viral peptides were not readily detected and are highly unlikely to be responsible for the accuracy of MALDI based SARS-CoV-2 diagnostics. Further investigation with more samples will be needed, but the methodology proves promising.

■ MATERIALS AND METHODS

Collection of Nasal Swab Specimens. Nasal swabs from the anterior nares were collected at the UC Davis Health Emergency Department (ED). The study was approved by the UC Davis Institutional Review Board. A subset of eight samples were selected for peptidomics in which half were positive and half were negative for COVID-19 (Table 1). COVID-19 diagnosis was cross-confirmed by United States Food and Drug Administration emergency use authorized molecular tests (digital droplet RT-PCR [Bio-Rad, Hercules, CA], and cobas Liat [Roche Diagnostics, Pleasanton, CA]). The age of participants ranges from 30 to 77 years. None of the patients were vaccinated. Three of the COVID-negative patients (n5, n6, and n9) had reported pre-existing pulmonary disease: two displayed chronic obstructive pulmonary disease (COPD) and one displayed moderate asthma with acute exacerbations.

Sample Preparation of Saline Media from Nasal Swabs. Endogenous peptides were processed by taking an aliquot of nasal swab transport media and separating the peptides from the remaining endogenous proteins and other large molecules by molecular weight cutoff using a 30 kDa centrifugal membrane filter (Amicon Ultra 0.5 mL, UFC503024, Sigma-Aldrich). Separated peptides were then assayed using a fluorescent peptide (PN 23290, Thermo Scientific) assay to determine total amount and analyzed by LC-MS/MS.

Liquid Chromatography Tandem Mass Spectrometry. LC peptide separation was done on a Dionex Ultimate RSLC (Thermo Scientific). The digested peptides were reconstituted in 0.1% trifluoroacetic acid, and 10 μL of each sample was loaded onto a PepMap C18 guard column: 100 μm × 2 cm, 5 μm particle size (PN 164564-CMD, Thermo Fisher), where they were desalted online before being separated on a PepMapRSLC C18 analytical column: 75 μm × 25 cm, 2
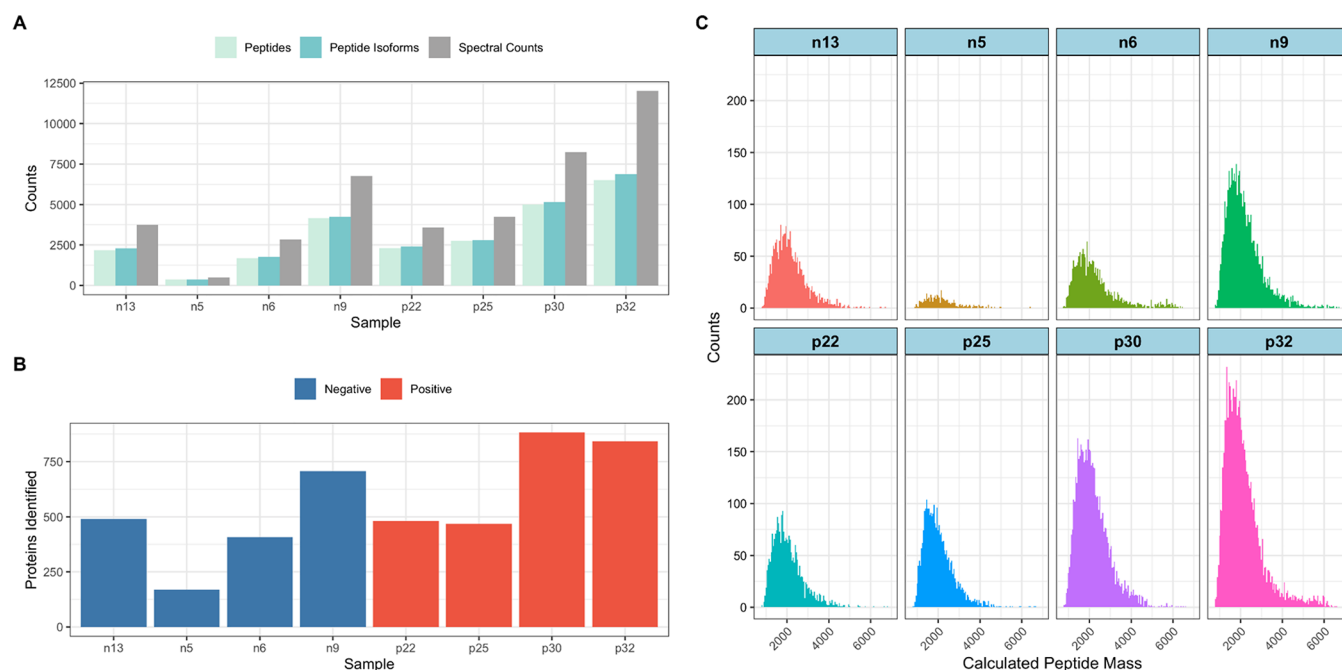
Figure 1. (A) Number of unique peptides, peptide isoforms, and spectral counts identified per sample. (B) Number of proteins identified per sample. (C) Histogram displaying the range and counts of masses generated per sample.

μm particle size (PN ES902, ThermoFisher). Peptides were eluted using a gradient of 0.1% formic acid (A) and 100% acetonitrile (B) with a flow rate of 300 nL/min. A 120 min gradient was run with 5% to 35% B over 50 min, 35% to 80% B over 3 min, 80% B for 1 min, 80% to 5% B over 1 min, and finally held at 5% B for 5 min.

Mass spectra were collected on an Orbitrap Fusion Lumos tribrid mass spectrometer (Thermo Fisher Scientific) in a data-dependent mode (Orbi/Orbi) with one MS precursor scan followed by 15 MS/MS scans. A dynamic exclusion of 35 s was used. MS spectra were acquired with a resolution of 70000 and a target of 1×10^6 ions or a maximum injection time of 20 ms. MS/MS spectra were acquired with a resolution of 17500 and a target of 5×10^4 ions or a maximum injection time of 250 ms. Peptide fragmentation was performed using higher-energy collision dissociation (HCD) with a normalized collision energy (NCE) value of 27. Unassigned charge states as well as +1 and ions greater than +5 were excluded from MS/MS fragmentation.

Data Analysis. Tandem mass spectra were searched using FragPipe, version 16.0 (MSFragger, version 3.3)[16] using the built in peptidomic workflow against combined the UniProt Human reference proteome (UP000005640_9606 20,588 entries), the UniProt Sars-CoV2 proteome (UP000464024, 17 entries), common laboratory contaminants, and an equal number of reverse decoy sequences. The search was performed twice: In one search for peptide identification, the peptide decoy false discovery rate (FDR) was set at 0.01 and protein decoy FDR was left open at 1. The second search was done using a traditional peptide and protein decoy FDR cutoff of 0.01 for protein identification. Output from FragPipe was analyzed using R. The primary outputs of interest from FragPipe used for our analysis are the combined_protein.tsv and the psm.tsv files from all the samples. The total FDR-filtered proteins from all experimental groups, in which each row is a protein group, is reported in from the combined_-protein.tsv file. The number of peptides found are from the

psm.tsv files. A separate psm.tsv was generated for each experiment and contains the FDR-filtered search results in which each row contains a peptide-spectrum match (PSM). For all files, the nonhuman entries were filtered out. To evaluate if a protein or peptide is present, we used the total number of PSMs with sequences mapping to the selected protein, including shared PSMs (Total Spectral Counts).

For the comparative analysis with the DIA study by Mun et al.,[17] we downloaded their supplementary file, pr1c00506_si_003.txt, and pulled the Protein_Accession column for comparisons. With the Human Atlas Protein, we pulled nasopharynx genes (https://www.proteinatlas.org/search/nasopharynx) on October 16, 2021 and used the Protein column for comparisons.

The top protein groups were ranked by spectral counts normalized by the length of protein with the highest value being the highest ranked. We selected the combined total spectral count (Combined_Total_Spectral_Count column) for normalization. The normalization reasoning is similar to iBAQ as longer proteins are expected to generate more peptides with proteolysis. This is approximated by dividing the spectral counts by the length of the protein. To generate the cumulative frequency graph, we then sorted the normalized spectral counts in descending order so that the protein with the highest total normalized spectral counts is the top rank. Last, we calculated the cumulative sum and divided each sum by the total. The top peptides were also identified using the same calculation and with spectral counts. For peptides, the spectral counts used are also the combined total because all psm.tsv files were concatenated, and the occurrence of each peptide was counted as a spectral count.

To identify potential proteases in the nasopharynx responsible for the endogenous peptides, we looked at peptides with at least one spectral count and generated a sequence motif for the preterminal, N-terminal, C-terminal, and post-terminal amino acids. The sequence motif was generated using the ggseqlogo R package.[18]
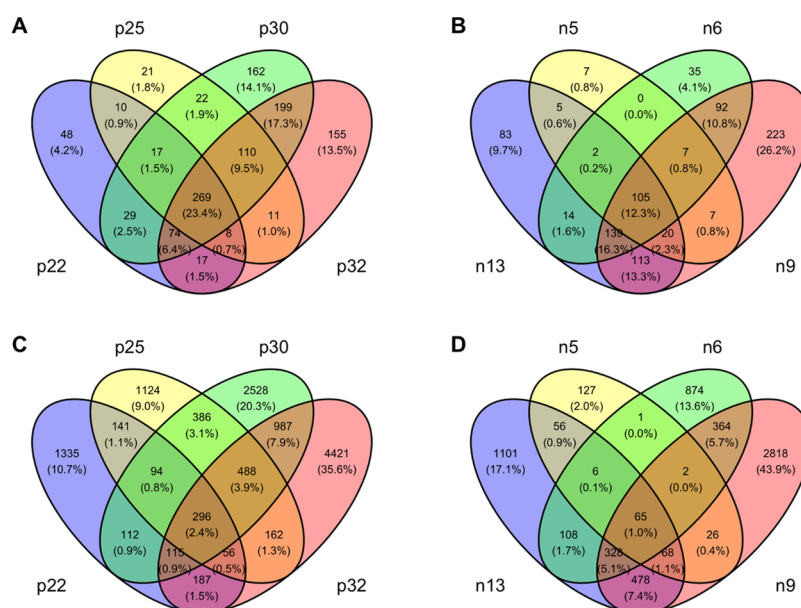
Figure 2. (A) Proteins in common between samples within the positive category. (B) Proteins in common between samples within the negative category. (C) Peptides in common between samples within the positive category. (D) Peptides in common between samples within the negative category.
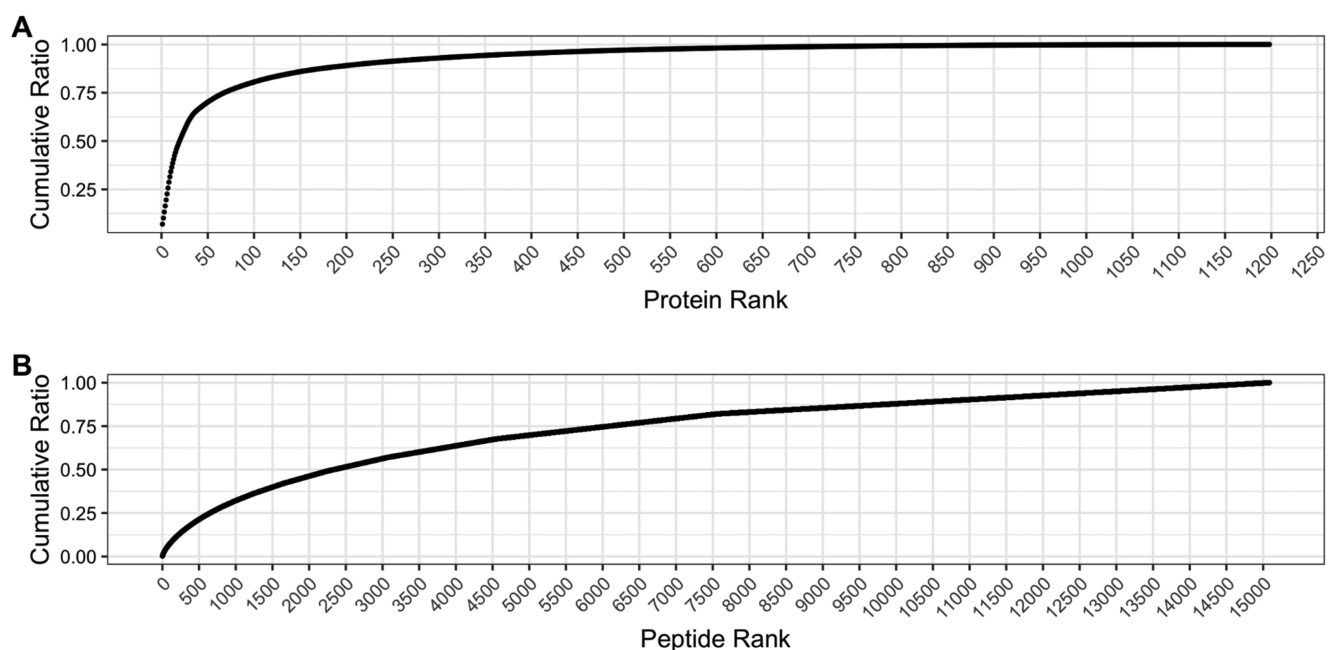

Figure 3. (A) Cumulative saturation graph for protein groups based on total spectral counts. (B) Cumulative saturation graph for peptides identified in our experiment.

To look for enriched pathways, we used Reactome and pulled 68 genes (this includes the indistinguishable mapped proteins) corresponding to the top 67 protein groups from the cumulative frequency analysis. The analysis was done on October 20, 2021 (https://reactome.org/userguide/analysis). The analysis included interactors.

Data Availability. All raw data and search results are available at the following repositories: Massive, https://massive.ucsd.edu/ (MSV000088411), and Proteome Exchange, http://proteomecentral.proteomexchange.org/ (PXD029800).

■ RESULTS AND DISCUSSION

Due to the high variability of both peptides and proteins identified between the nasal swabs and the low power of this study ($n = 4$), we did not test for the differentiation of proteins and peptides between positive and negative cohorts (Figure 1).

Nevertheless, we identified 14270 endogenous peptides across 1198 protein groups that we hypothesize could be partly responsible for the previously reported MALDI-TOF-MS based screen.[6] Peptides can exist in different isoforms due to post-translational modifications such as N-terminal acetylation and deamidation. These modifications can have real biological significance and can also be introduced during the preparation

Table 2. Top 20 Protein Groups Based on Total Combined Spectral Counts

rank	cumulative frequency	protein ID	gene	protein length	coverage	description	protein probability	top peptide probability	combined total peptides	combined spectral count	combined unique spectral count	combined total spectral count	count mapped proteins
1	0.07	P02808	STATH	62	75.6	statherin	1.00	1.00	181	1121	1121	1121	1
2	0.10	P62807	H2BC4	126	99.2	histone H2B type 1-C/E/F/G/I	0.26	1.00	212	1017	0	1018	1
3	0.13	Q5QNW6	H2BC18	126	99.2	histone H2B type 2-F	1.00	1.00	210	7	7	1017	1
4	0.16	P58876	H2BC5	126	99.2	histone H2B type 1-D	1.00	1.00	209	6	6	1017	1
5	0.20	Q99877	H2BC15	126	99.2	histone H2B type 1-N	1.00	1.00	209	3	3	1014	1
6	0.23	Q99879	H2BC14	126	88.9	histone H2B type 1-M	1.00	1.00	208	3	3	1013	1
7	0.26	Q16778	H2BC21	126	99.2	histone H2B type 2-E	1.00	1.00	207	50	0	993	1
8	0.29	O60814	H2BC12	126	99.2	histone H2B type 1-K	1.00	1.00	200	14	0	987	1
9	0.32	Q99880	H2BC13	126	99.2	histone H2Btype 1-L	1.00	1.00	174	2	2	905	1
10	0.34	P60709	ACTB	375	92.3	actin cytoplasmic 1	1.00	1.00	460	2571	3	2571	1
11	0.36	Q8FI13	H2AC18	130	99.2	histone H2A type 2-A	1.00	1.00	174	742	0	742	1
12	0.39	P0C0S8	H2AC11	130	99.2	histone H2A type 1	1.00	1.00	167	29	0	721	1
13	0.40	P62328	TMSB4X	44	97.7	thymosin β-4	1.00	1.00	33	220	220	220	1
14	0.42	P62805	H4C1	103	95.1	histone H4	1.00	1.00	142	512	512	512	1
15	0.44	Q93077	H2AC6	130	99.2	histone H2A type 1-C	1.00	1.00	149	159	0	560	1
16	0.46	P01833	PIGR	764	48.4	polymeric immunoglobulin receptor	1.00	1.00	364	3087	3087	3087	1
17	0.47	Q71DI3	H3C15	136	83.8	histone H3.2	1.00	1.00	161	473	13	473	1
18	0.48	P04406	GAPDH	335	97.6	glyceraldehyde-3-phosphate dehydrogenase	1.00	1.00	311	1065	1062	1065	1
19	0.49	Q16695	H3-4	136	86.8	histone H3.1t	1.00	1.00	141	120	120	432	1
20	0.51	Q8IUE6	H2AC21	130	87.7	histone H2A type 2-B	1.00	1.00	122	42	17	393	1

Table 3. Top 20 Peptides Based on Total Spectral Counts

rank	cumulative frequency	peptide	pre-amino acid	post-amino acid	peptide length	calculated peptide mass	protein ID	gene	description	count mapped proteins	combined total spectral count
1	0.002435181	SLAKADAAP	V	A	24	2602.324	P01833	PIGR	polymeric immunoglo-bulin receptor	1	102
2	0.004106384	AVEERKAAG	V	A	37	3971.035	P01833	PIGR	polymeric immunoglo-bulin receptor	1	70
3	0.005753712	ADKPDMGEI	M	-	43	4949.518	P63313	TMSB10	thymosin β-10	1	69
4	0.007257795	SLAKADAAP	V	L	30	3282.684	P01833	PIGR	polymeric immunoglo-bulin receptor	1	63
5	0.008666380	AVADTRDQA	K	L	39	3831.847	P01833	PIGR	polymeric immunoglo-bulin receptor	1	59
6	0.009955594	PDEKVLDSG	A	A	16	1874.937	P01833	PIGR	polymeric immunoglo-bulin receptor	1	54
7	0.011125435	DVSLAKADA	R	A	26	2816.419	P01833	PIGR	polymeric immunoglo-bulin receptor	1	49
8	0.012223655	AVEERKAAG	V	K	36	3842.940	P01833	PIGR	polymeric immunoglo-bulin receptor	1	46
9	0.013274125	EIENKAIQDP	R	A	30	3371.670	P01833	PIGR	polymeric immunoglo-bulin receptor	1	44
10	0.014300721	AKADAAPDE	L	A	22	2402.207	P01833	PIGR	polymeric immunoglo-bulin receptor	1	43
11	0.015303443	LFAEEKAVAI	R	S	38	3881.826	P01833	PIGR	polymeric immunoglo-bulin receptor	1	42
12	0.016282290	SLAKADAAP	V	K	23	2474.229	P01833	PIGR	polymeric immunoglo-bulin receptor	1	41
13	0.017261137	VESTGVFTTF	V	S	26	2715.438	P04406	GAPDH	glyceraldehyde-3-phos-phate dehydrogen-ase	1	41
14	0.01B216110	PPAGQPOGI	A		23	2355.231	P04280	PRB1	basic salivary proline-rich protein 1	2	40
15	0.019147209	AIQDPRLFAE	K	A	41	4278.998	P01833	PIGR	polymeric immunoglo-bulin receptor	1	39
16	0.020054433	ELRVAPEEHI	N	M	30	3438.829	P60709	ACTB	actin, cytoplasmic 1	2	38
17	0.020937784	AIQDPRLFAE	K	L	42	4350.035	P01833	PIGR	polymeric immunoglo-bulin receptor	1	37
18	0.021797259	AVVVKKIETR	R		24	2596.480	P05787	KRT8	keratin, type II cytoskeletal 8	1	36
19	0.022656735	EERKAAGSR	V	A	35	3800.929	P01833	PIGR	polymeric immunoglo-bulin receptor	1	36
20	0.023516211	EIENKAIQDP	R	D	23	2642.366	P01833	PIGR	polymeric immunoglo-bulin receptor	1	36

of samples. With our analysis, we identified 15086 unique peptide isoforms. We identified 96 proteins in which all samples have at least one spectral count and 44 peptides in which all samples have the presence of at least one spectral count. Lowering the threshold in which seven or more samples have the presence of at least one spectral count, we identified 196 proteins. Within confirmed COVID-19 positive and negative samples, we identified 269 protein groups that all positive samples have in common and 105 protein groups that all negative samples have in common (Figure 2A and 2B). For peptides, we identified 296 common peptides and 65 common peptides within the positive and negative categories, respectively (Figure 2C,D). We identified three proteins that are uniquely found in the positive samples (ANXA5, CANX, SCFD1) and no proteins unique to negative samples. We identified six peptides unique to the positive samples and one peptide unique to the negative samples.

To identify the peptides and proteins in these samples that were the most highly abundant, we identified 67 protein groups (out of 1198) that had the highest number of peptides (normalized by protein length) and cumulatively account for 75% of the total peptides found in this experiment (Figure 3A and Supplementary Table 1). Of these protein groups, the top 20 are listed in Table 2. We also identified 6093 peptides that had the highest number of spectral counts and cumulatively

account for 75% of the total peptides (Figure 3B and Supplementary Table 2). These peptides correspond to 1015 proteins, and the summary of counts can be found in Supplementary Table 3. Of these peptides, the top 20 are listed in Table 3.

Although it is tempting to match the m/z values of the peptides identified in this study with values reported previously in the MALDI-TOF-MS based assay, matching such data would be an educated guess at best. First, there are inherent differences between LC-MS/MS and MALDI-TOF-MS, including ionization, peptide suppression, matrix effects, and the lack of isotopic resolution in the MALDI-TOF-MS due to data smoothing. Second, our LC-MS/MS analysis in this study should be far more sensitive than the MALDI-TOF-MS based assay. However, it is a reasonable hypothesis that peptides identified in this study by LC-MS/MS are responsible for some of the m/z values seen in our previous MALDI-TOF-MS based assay.

Comparing the masses (Da) of the peptides identified by LC-MS/MS and m/z values of the MALDI-TOF-MS assay, we found that the ranges overlap but do not align perfectly and the number of masses identified by LC-MS/MS is far greater than the MILO curated number from MALDI-TOF-MS (Figure 4). The mass range of peptides identified is between 768.4 and 6941.4 Da (Da), with a mean of 2144.3 Da and a median of
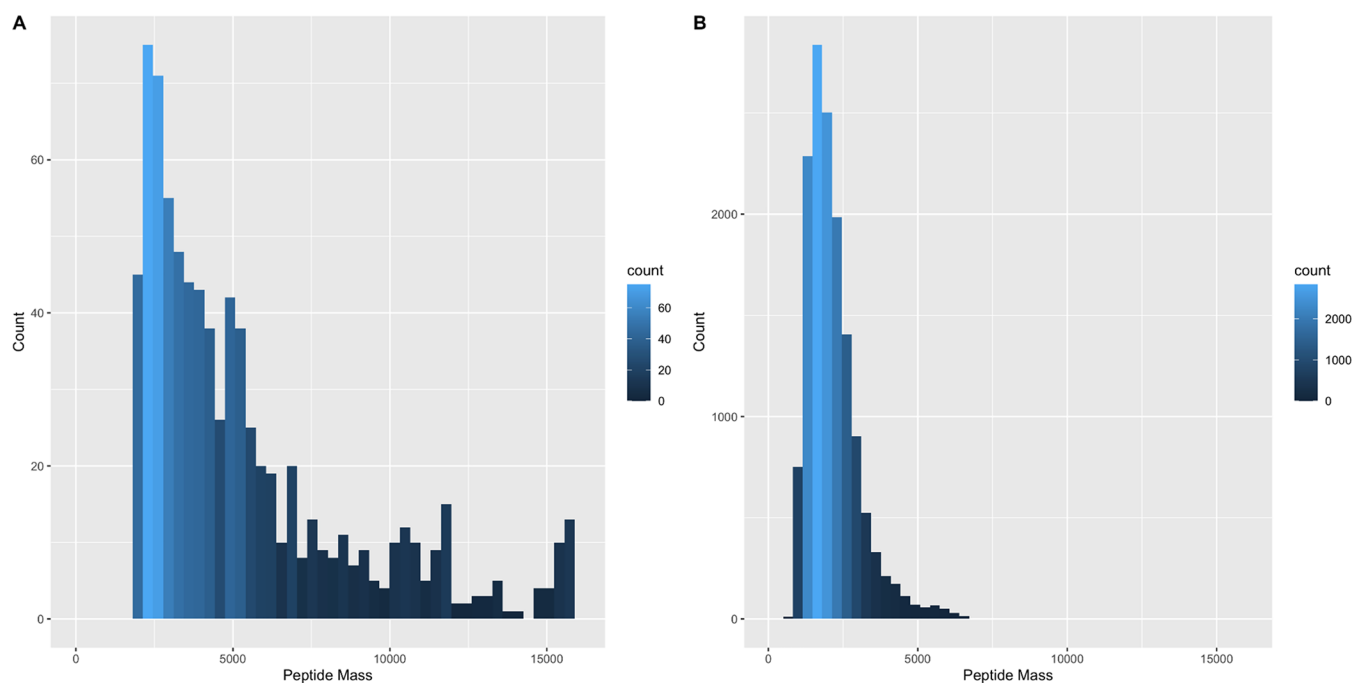
Figure 4. (A) Mass distribution of the 816 MILO predicted masses determined to be significant, ranging from 1992.7 to 16019.0 m/z. (B) Mass distribution of the peptides identified by LC-MS/MS, 14318 unique masses, ranging from 768.4 to 6941.4 Da.

Table 4. Distribution of Mass Ranges by Sample

sample	count	min	1st qu.	median	mean	3rd qu.	max
n13	2282	821.44	1611.85	2029.05	2172.07	2563.24	6693.10
n5	359	920.51	1591.81	2032.05	2196.30	2551.77	6394.22
n6	1771	768.39	1554.40	2021.01	2280.75	2627.80	6452.18
n9	4240	792.41	1571.81	1987.62	2170.46	2552.79	6707.25
p22	2401	821.44	1525.75	1906.96	2057.89	2427.32	6824.06
p25	2788	821.44	1499.79	1886.96	2024.19	2381.58	6669.05
p30	5160	818.49	1600.78	2031.03	2173.90	2579.70	6587.68
p32	6860	799.44	1498.83	1924.01	2137.86	2492.29	6941.44

1965.1 Da. The range of masses for each sample can be found in Table 4. The MALDI-TOF-MS m/z range is between 1992.7 and 16019.0, with a mean of 5601 m/z and a median of 4307 m/z.

Although it is likely that the molecules detected in the MALDI-TOF-MS based assay are composed mostly of human host response proteins and peptides, it does not rule out the possibility that other molecules such as lipids and carbohydrates not detected in this study may be responsible in part for the MALDI-TOF-MS assay's performance.

Of the peptides identified in our study, none corresponded with SARS-CoV-2 viral proteins. In subsequent experiments, viral proteins were detected on nasal swabs using traditional bottom-up proteomics and were relatively low in abundance compared to human host proteins (data not shown). Using a diaPASEF analysis like Mun et al.,[17] viral proteins were 100−1000 times less abundant than the most abundant human host proteins detected (complete data reported in subsequent publication). Bottom-up proteomics assays, where the proteins are digested using a protease and then detected, are far more sensitive than the native peptidomic workflow presented here. This decrease in sensitivity is due mainly to the massively expanded search space of nonenzymatic peptidomic searches when combined with decoy false discovery filtering.

The human protein groups identified in this study generally matched the proteins expected to be in the nasopharynx. The Human Protein Atlas lists 365 genes reported to be in the nasopharynx (https://www.proteinatlas.org/search/nasopharynx). Of that, we found 35 proteins (Supplementary Table 4). Compared with a previous bottom-up proteomics analysis of nasal swabs, our results are generally consistent. In a recent DIA-based bottom-up proteome profiling of nasopharyngeal swabs, Mun et al.[17] reported 7674 proteins identified. We analyzed the protein groups from their list of detected proteins using the Spectronaut results from their published repository (PXD025277). From that, we extracted 7805 protein identifications in 7711 protein groups. In this study, 90% of the proteins we identified (1116 of 1245) matched the data in their bottom-up DIA study (Supplementary Table 5).

Analyzing the protease cut sites of the peptides, we identified neutrophil elastase (P08246) as a possible protease in the nasopharynx responsible for the endogenous peptides. There is a high number of valines in the preterminal amino acid position, which is a known specificity for this enzyme (Figure 5A). The peptide coverage of the protease was high, 22.8%, and we found spectral counts for this protein in seven of eight samples. The sample in which the neutrophil elastase was not detected was n5, which is the sample with the lowest number of spectral counts. For this protein, there are 35 combined
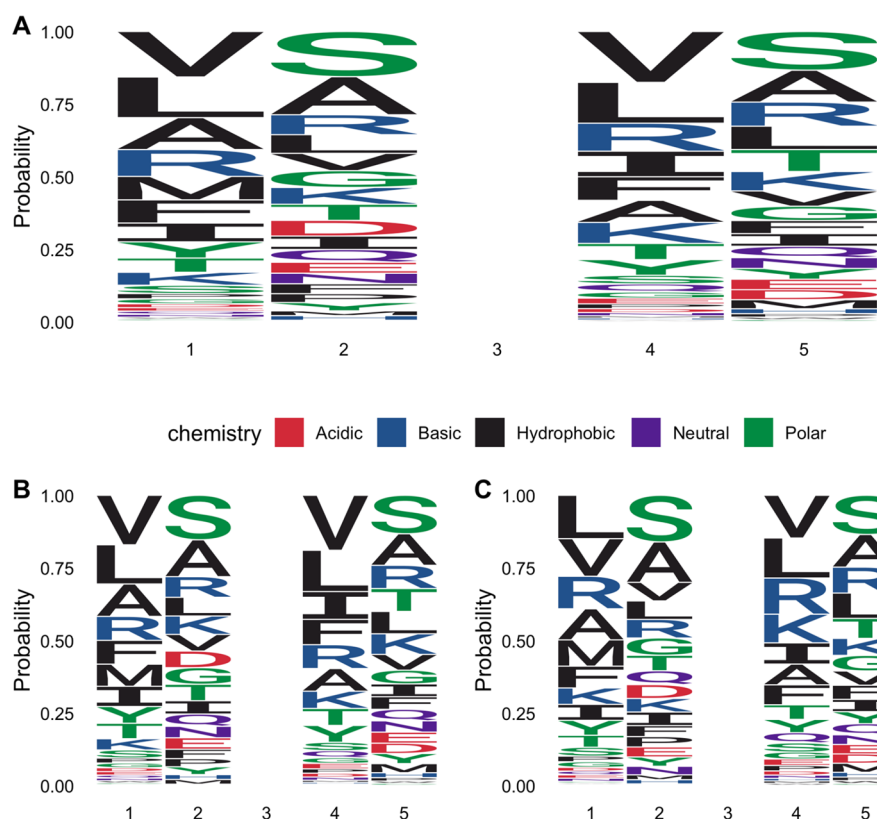
Figure 5. Sequence motif for the preterminal (1), N-terminal (2), C-terminal (4), and post-terminal amino acids (5). (A) Sequence motif for all samples. (B) Sequence motif for positive samples. (C) Sequence motif for negative samples.
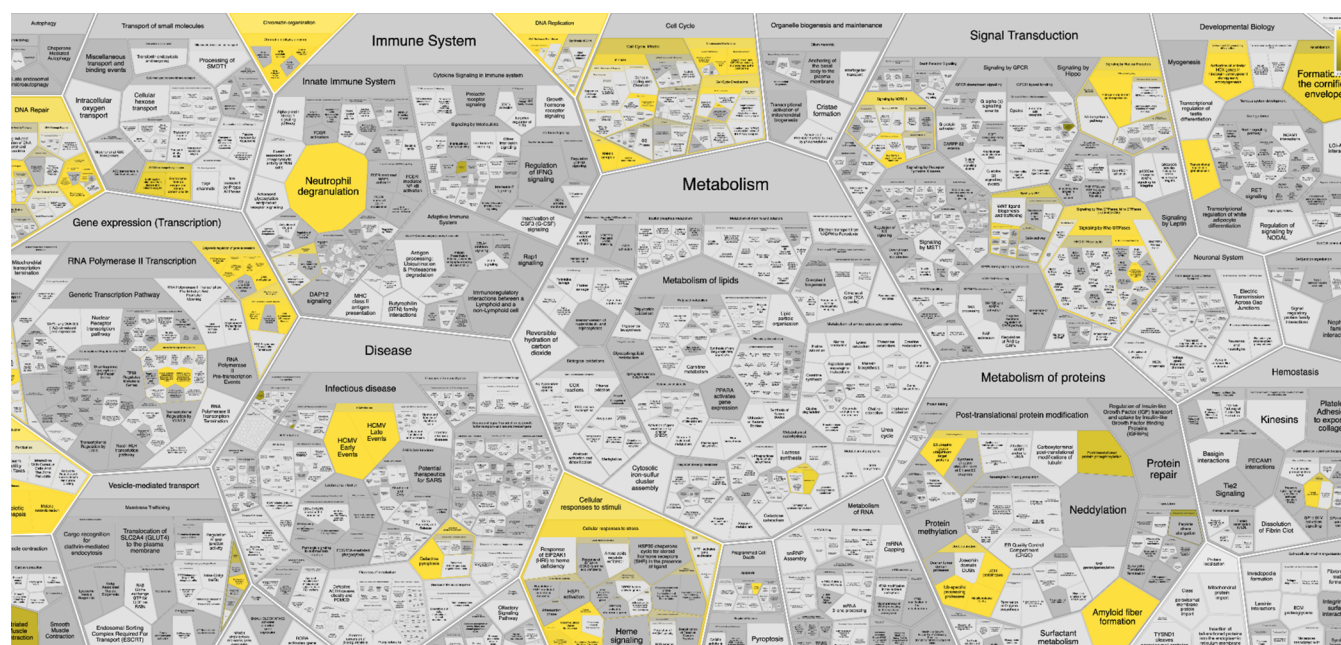

Figure 6. Reacfoam output from Reactome, a holistic view of all the human pathways. The scale to the top right indicates the *p*-value obtained from over-representation analysis for molecules selected for each pathway result. The top of the scale in yellow is near zero, and the bottom with grayish yellow is the 0.05 threshold.

spectral counts (razor), 34 combined unique spectral counts and 35 combined total spectral counts. The sequence motif between the positive and negative samples do not appear to be significantly different with the top amino acids changing only slightly (Figure 5B).

Selecting the genes from the top 67 protein groups in our top protein cumulative frequency analysis (68 genes including the indistinguishable mapped proteins), we looked for enriched pathways using Reactome (Figure 6). Of the 68 genes, four were not found. The top five pathways found are

involved in DNA methylation, packaging of telomere ends, methylation of histones and DNA by Polycomb Repressive Complex 2 (PRC2), deacetylation of histones by histone deacetylases (HDACs), and nucleosome assembly (complete list available in Supplementary Table 6).

CONCLUSIONS

Using our peptidomic workflow, we identified 14270 endogenous peptides across 1245 protein groups from nasal swab transport media. The proteins mapped to these peptides are primarily polymeric immunoglobulin receptor, actin, statherin, glyceraldehyde-3-phosphate dehydrogenase, thymosin β-4, and histones. Our method identified protease cut sites but was not sensitive enough to detect SARS-CoV-2 viral peptides. Due to the large biological diversity typically seen in studies like this, a larger number of samples will be needed to validate these results. We believe that the result from our methodology is promising and that some of the peptides seen in this limited sample set should be representative of the m/z signals seen in our previous MALDI-TOF assay.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c01864.

> MSFragger parameters datasheet (TXT)
>
> Supplementary Table 1, protein group cumulative saturation (TXT)
>
> Supplementary Table 2, peptide cumulative saturation (TXT)
>
> Supplementary Table 3, summary of the peptide cumulative saturation table in which proteins identified are grouped and the number of peptides within the group are tallied (TXT)
>
> Supplementary Table 4, Human Protein Atlas Hits (TXT)
>
> Supplementary Table 5, Mun 2021 IDA Hits (TXT)
>
> Supplementary Table 6, complete list of enriched pathways found by Reactome (TXT)

AUTHOR INFORMATION

Corresponding Author

Brett S. Phinney − Proteomics Core, University of California, Davis, Davis, California 95616, United States; orcid.org/0000-0003-3870-3302; Email: bsphinney@ucdavis.edu

Authors

Helen Tsai − Proteomics Core, University of California, Davis, Davis, California 95616, United States; orcid.org/0000-0002-4043-8993

Gabriela Grigorean − Proteomics Core, University of California, Davis, Davis, California 95616, United States

Michelle R. Salemi − Proteomics Core, University of California, Davis, Davis, California 95616, United States

Hooman H. Rashidi − Department of Pathology and Laboratory Medicine, University of California, Davis, Sacramento, California 95817, United States

John Pepper − SpectraPass, LLC, Las Vegas, Nevada 89135, United States; Allegiant Air, Las Vegas, Nevada 89144, United States

Nam K. Tran − Department of Pathology and Laboratory Medicine, University of California, Davis, Sacramento, California 95817, United States

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.2c01864

ACKNOWLEDGMENTS

REFERENCES

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

(2) Kipping, M.; Tänzler, D.; Sinz, A. A Rapid and Reliable Liquid Chromatography/mass Spectrometry Method for SARS-CoV-2 Analysis from Gargle Solutions and Saliva. *Anal. Bioanal. Chem.* 2021, *413*, 6503.

(3) 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.; Mangalaparthi, K. K.; et al. A Mass Spectrometry-Based Targeted Assay for Detection of SARS-CoV-2 Antigen from Clinical Specimens. *EBioMedicine* 2021, *69*, 103465.

(4) St-Germain, J. R.; Astori, A.; Raught, B. A SARS-CoV-2 Peptide Spectral Library Enables Rapid, Sensitive Identification of Virus Peptides in Complex Biological Samples. *J. Proteome Res.* 2021, *20*, 2187−2194.

(5) Iles, R. K.; Zmuidinaite, R.; Iles, J. K.; Carnell, G.; Sampson, A.; Heeney, J. L. Development of a Clinical MALDI-ToF Mass Spectrometry Assay for SARS-CoV-2: Rational Design and Multi-Disciplinary Team Work. *Diagnostics (Basel)* 2020, *10*, 746.

(6) Tran, N. K.; Howard, T.; Walsh, R.; Pepper, J.; Loegering, J.; Phinney, B.; Salemi, M. R.; Rashidi, H. H. Novel Application of Automated Machine Learning with MALDI-TOF-MS for Rapid High-Throughput Screening of COVID-19: A Proof of Concept. *Sci. Rep.* 2021, *11*, 8219.

(7) Rocca, M. F.; Zintgraff, J. C.; Dattero, M. E.; Santos, L. S.; Ledesma, M.; Vay, C.; Prieto, M.; Benedetti, E.; Avaro, M.; Russo, M.; et al. A Combined Approach of MALDI-TOF Mass Spectrometry and Multivariate Analysis as a Potential Tool for the Detection of SARS-CoV-2 Virus in Nasopharyngeal Swabs. *J. Virol. Methods* 2020, *286*, 113991.

(8) Nachtigall, F. M.; Pereira, A.; Trofymchuk, O. S.; Santos, L. S. Detection of SARS-CoV-2 in Nasal Swabs Using MALDI-MS. *Nat. Biotechnol.* **2020**, *38*, 1168−1173.

(9) Lazari, L. C.; Ghilardi, F. D. R.; Rosa-Fernandes, L.; Assis, D. M.; Nicolau, J. C.; Santiago, V. F.; Dalçóquio, T. F.; Angeli, C. B.; Bertolin, A. J.; Marinho, C. R. Prognostic Accuracy of MALDI-TOF Mass Spectrometric Analysis of Plasma in COVID-19. *Life Sci. Alliance* **2021**, *4*, e202000946.

(10) Pranada, A. B.; Schwarz, G.; Kostrzewa, M. MALDI Biotyping for Microorganism Identification in Clinical Microbiology. In *Advances in MALDI and Laser-Induced Soft Ionization Mass Spectrometry*; Cramer, R., Ed.; Springer International Publishing: Cham, 2016; pp 197−225.

(11) Nomura, F. Proteome-Based Bacterial Identification Using Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS): A Revolutionary Shift in Clinical Diagnostic Microbiology. *Biochim. Biophys. Acta* **2015**, *1854*, 528−537.

(12) Croxatto, A.; Prod'hom, G.; Greub, G. Applications of MALDI-TOF Mass Spectrometry in Clinical Diagnostic Microbiology. *FEMS Microbiol. Rev.* **2012**, *36*, 380−407.

(13) Tsuchida, S. Application of MALDI-TOF for Bacterial Identification. *Use of Mass Spectrometry Technology (MALDI-TOF) in Clinical Microbiology* **2018**, 101−112.

(14) Swiatly, A.; Horala, A.; Hajduk, J.; Matysiak, J.; Nowak-Markwitz, E.; Kokot, Z. J. MALDI-TOF-MS Analysis in Discovery and Identification of Serum Proteomic Patterns of Ovarian Cancer. *BMC Cancer* **2017**, *17*, 472.

(15) Böhm, D.; Keller, K.; Pieter, J.; Boehm, N.; Wolters, D.; Siggelkow, W.; Lebrecht, A.; Schmidt, M.; Kölbl, H.; Pfeiffer, N.; et al. Comparison of Tear Protein Levels in Breast Cancer Patients and Healthy Controls Using a de Novo Proteomic Approach. *Oncol. Rep.* **2012**, *28*, 429−438.

(16) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry-based Proteomics. *Nat. Methods* **2017**, *14*, 513−520.

(17) Mun, D.-G.; Vanderboom, P. M.; Madugundu, A. K.; Garapati, K.; Chavan, S.; Peterson, J. A.; Saraswat, M.; Pandey, A. DIA-Based Proteome Profiling of Nasopharyngeal Swabs from COVID-19 Patients. *J. Proteome Res.* **2021**, *20*, 4165−4175.

(18) Wagih, O. Ggseqlogo: A Versatile R Package for Drawing Sequence Logos. *Bioinformatics* **2017**, *33*, 3645−3647.