

Rapid Detection of COVID-19 Using MALDI-TOF-Based Serum Peptidome Profiling

Ling Yan,[▽] Jia Yi,[▽] Changwu Huang,[▽] Jian Zhang,[▽] Shuhui Fu, Zhijie Li, Qian Lyu, Yuan Xu, Kun Wang, Huan Yang, Qingwei Ma, Xiaoping Cui, Liang Qiao,* Wei Sun,* and Pu Liao*



Cite This: <https://dx.doi.org/10.1021/acs.analchem.0c04590>



Read Online

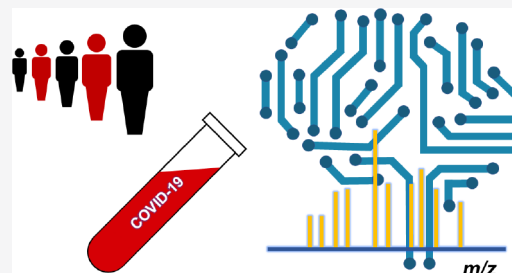
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The outbreak of coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 is ongoing and a serious threat to global public health. It is essential to detect the disease quickly and immediately to isolate the infected individuals. Nevertheless, the current widely used PCR and immunoassay-based methods suffer from false negative results and delays in diagnosis. Herein, a high-throughput serum peptidome profiling method based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is developed for efficient detection of COVID-19. We analyzed the serum samples from 146 COVID-19 patients and 152 control cases (including 73 non-COVID-19 patients with similar clinical symptoms, 33 tuberculosis patients, and 46 healthy individuals). After MS data processing and feature selection, eight machine learning methods were used to build classification models. A logistic regression machine learning model with 25 feature peaks achieved the highest accuracy (99%), with sensitivity of 98% and specificity of 100%, for the detection of COVID-19. This result demonstrated a great potential of the method for screening, routine surveillance, and diagnosis of COVID-19 in large populations, which is an important part of the pandemic control.



Coronavirus disease-2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has become an unprecedented global threat^{1,2} and was declared a pandemic by the World Health Organization on March 11, 2020.³ Accurate, rapid and high-throughput detection of COVID-19 patients, especially among those with similar symptoms, is very important for the epidemic control.

A number of COVID-19 identification methods based on the detection of virus nucleic acids, virus proteins, or specific antibodies generated by the human immune system have been developed.⁴ Reverse transcription polymerase chain reaction (RT-PCR) is the most predominantly used method for the detection of SARS-CoV-2 nucleic acids using respiratory samples⁵ and has become the standard clinical method for the diagnosis of COVID-19. However, improper sampling operations, low viral load in the sample collection area or at the collection time, nonstandard transportation processes, and viral mutations often lead to false negative results.^{6,7} The test sensitivity of RT-PCR for throat swab samples has been reported to be only from 30% to 60%,^{6,8} and multiple evidences are recommended for epidemiologically linked patients even when the nasopharyngeal and/or oropharyngeal swab tests are negative.⁷ New techniques, such as CARMEN-Cas13⁹ and isothermal amplification,¹⁰ have also emerged. In addition, mass spectrometry-based methods have also been used to detect unique peptides of SARS-CoV-2 nucleoproteins¹¹ and nucleocapsid N protein¹² in the physiological fluids

of infected patients. Immunoassays for the detection of specific immunoglobulin G and M (IgG and IgM) corresponding to SARS-CoV-2 virus infection in serum are alternative methods of clinical diagnosis of COVID-19. However, there is a distinct delay for the generation of the antibodies after infection. It is reported that the median interval for the generation of SARS-CoV-2 specific IgG and IgM is 13 days post symptom onset, and it takes approximately 17–19 days or 20–22 days after symptom onset that positive rates for IgG reach 100% or positive rates for IgM reach 94.1%, respectively.¹³ Chest CT has also been used in clinical diagnosis of SARS-CoV-2 infection. It is a noninvasive imaging examination with high speed and high test sensitivity (86%–98%),^{6,14–16} but the test specificity is very low (25%)⁶ with a relatively high cost and requirement for experienced experts. To date, there is still an urgent need of accurate and high-throughput detection of COVID-19 for large population screening.

Besides targeting virus nucleic acids, virus proteins, or infection specific antibodies, detection of the induced molecular changes in the host during and after infection is

Received: October 30, 2020

Accepted: February 22, 2021

another effective way for disease diagnosis and deciphering the host response to virus. For example, elevation of the C-reactive protein and D-dimer as well as a decrease of lymphocytes, leukocytes, and blood platelets have been reported in infected patients,¹⁴ and cytokines IP-10 and MCP-3 were found as severity predictors for the progression of COVID-19.¹⁷ Human serum contains a complex array of proteolytically derived peptides, namely, serum peptidome, that is correlated with the biological events occurring in the entire organism.^{18,19} Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)-based peptidome analysis, as a rapid, high-throughput, low cost, and easily handled technique, has been successfully used in clinic applications, including microorganism identification²⁰ approved by the U.S. Food and Drug Administration, and determining treatment benefits from epidermal growth factor receptor (EGFR) inhibitors in nonsmall cell lung cancer patients as a laboratory-developed test in clinics.²¹ It has also shown great potential for the screening of Alzheimer's disease²² and rapid diagnosis and treatment monitoring of active *Mycobacterium tuberculosis* (TB) disease.²³

In this study, serum samples of 146 COVID-19 patients, including mild, typical, severe, and critical classifications, and 152 control individuals, including non-COVID-19 patients with fever/cough symptoms, TB patients, and healthy controls, were analyzed by MALDI-TOF MS. Twenty-five MS peaks with statistically significant differences between the patients and control individuals were identified. Using various machine learning methods, a model constructed with the 25 peaks was established, showing high accuracy (99%) for the identification of COVID-19 patients with a sensitivity of 98% and specificity of 100% on a test cohort (100 samples) independent from the samples used for model generation and feature selection. This accurate and rapid method provides a powerful tool for high-throughput screening and surveillance of COVID-19.

EXPERIMENTAL SECTION

Patients and Samples. Non-COVID-19 serum samples were collected as the control, including 73 samples from non-COVID-19 patients with similar clinical symptoms, 33 samples from TB patients, and 46 samples from healthy individuals. These samples were collected from Chongqing General Hospital. The serum samples from 146 COVID-19 patients were collected from Sanxia Hospital. All the patients were diagnosed as COVID-19 according to the Chinese Government Diagnosis and Treatment Guideline. The COVID-19 virus nucleic acid detections in throat swabs were all positive. The computed tomography (CT) ground-glass opacity pneumonia characteristic was detected in 97% (142/146) of the patients at the sampling time. The symptoms, such as fever, cough, fatigue, and dyspnea, severity classification, and other information are shown in Table S1 and Data set S1. Classification of COVID-19 patients was performed according to the following guideline: (1) Mild ($n = 2$): mild symptoms without pneumonia, (2) Typical ($n = 114$): fever or respiratory tract symptoms with pneumonia, (3) Severe (fulfill any of the three criteria, $n = 26$): respiratory distress with respiratory rate ≥ 30 times/min, oxygen saturation $\leq 93\%$ in resting state, arterial blood oxygen partial pressure (PaO_2) /oxygen concentration (FiO_2) ≤ 300 mmHg, (4) Critical (fulfill any of the three criteria, $n = 4$): respiratory failure and require mechanical ventilation, shock incidence, admission to ICU with other organ failure. The peripheral venous blood was

collected in the morning before eating using a vacuum serum tube without additives (Shandong Weihai Weigao Blood Collection Supplies Co., Ltd., China). After centrifugation at 2264g for 10 min and 56 °C sterilization for 30 min, the serum samples were aliquoted and frozen at -80 °C.

Sample Preparation and MALDI-TOF Analysis. For each sample, 5 μL of serum was diluted with 45 μL of dilution buffer (PMFpre kit 1010305, Bioyong Technologies Inc., Beijing, China). Then, 10 μL of the diluted serum was mixed with 10 μL of sinapinic acid matrix solution, and 2 μL of the mixture was pipetted onto a stainless-steel target plate. After being dried at room temperature, the samples were analyzed on a MALDI-TOF MS (Clin-TOF-II; Bioyong Technologies Inc., Beijing, China) in a linear positive mode. Mass calibration was performed with the standard calibration mixture of peptides and proteins (Sigma-Aldrich Co., St. Louis, MO, USA) with a tolerance of 500 ppm. Each spectrum was accumulated with 500 laser shots (50 positions per sample spot and 10 laser shots per position).

MALDI-TOF Data Analysis. MALDI-TOF raw data were processed with R packages MALDIquant, MALDIquantForeign,²⁴ and limma.²⁵ The data of COVID-19 patients and control samples were randomly split into training and validation cohorts with an allocation of 2:1, corresponding to 198 (101 controls and 97 patients) and 100 (51 controls and 49 patients) samples, respectively. The partial least-squares-discriminant analysis (PLS-DA) was performed using Metaboanalyst 4.0²⁶ (McGill University, Montreal, Canada, <https://www.metaboanalyst.ca/>). A two-tailed t test was used to identify the differentially expressed peaks between COVID-19 and control samples on the training data set. The P values were corrected by the Benjamini–Hochberg algorithm. Peaks with adjusted $P < 0.05$ and fold change greater than 1.5 were used for further feature selection with the least absolute shrinkage and selection operator (LASSO)²⁷ regression conducted in R version 3.5.3 with the glmnet²⁸ package. Recursive feature elimination with cross-validation (RFECV) was performed using the sklearn.feature_selection.RFECV package in Jupyter Notebook 5.7.9. The estimator of parameters was logical regression, and the CV of parameters was 10.²⁹ Logistic regression (LR),³⁰ support vector machine (SVM),³¹ naive bayes (NB),³² decision tree (DT),³³ random forest (RF),³⁴ gradient boosting decision tree (GBDT),^{35,36} AdaBoost,³⁷ and K-nearest neighbors (KNN)³⁸ were performed using the sklearn package in Jupyter Notebook 5.7.9. Detailed information on data analysis is shown in Text S1.

Ethical Statement. This study was approved by the Ethical Committee of Chongqing General Hospital. Informed consents from patients were waived by the board. The samples in this study were from a clinical trial registered in the Chinese Clinical Trial Registry (ChiCTR2000033872).

RESULTS AND DISCUSSION

Selection of Feature Peaks on MALDI-TOF Mass Spectra. The workflow of this study is shown in Figure 1. After sterilization, the serum samples were profiled by a MALDI-TOF MS. The obtained mass spectra were then subjected to peak extraction and alignment, followed by feature selection and classification model construction by using different machine learning methods. The representative mass spectra of samples in different groups are shown in Figure S1. Most of the MS peaks were located between m/z 5000 and m/z 30,000. Distinctive features can be easily observed from the

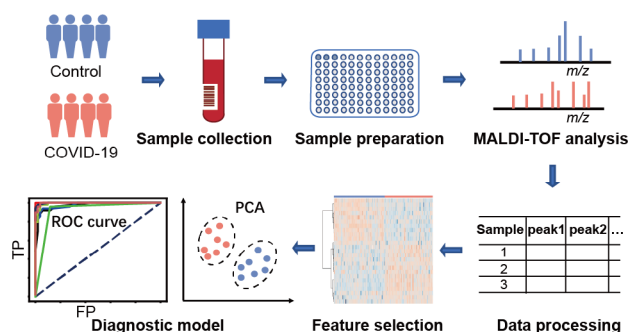


Figure 1. Scheme of establishing a diagnostic model for rapid screening of COVID-19 patients. Serum samples collected from COVID-19 patients and control participants were analyzed with MALDI-TOF after simple pretreatment. Mass spectra were aligned with MALDIquant, and significant features were selected to establish the diagnostic model with different machine learning methods.

profiles, e.g., $m/z = 6357, 6654, 6639, 13,886,$ and $28,232$ that were significantly down-regulated in the COVID-19 group, and $m/z = 7614, 15,123, 15,867,$ and $28,091$ that were significantly up-regulated in the COVID-19 group. Totally, 386 peaks were detected after data processing of all the MS raw files (Data set S2).

Feature peaks were selected according to the process shown in Figure 2a. After intensity normalization and missing value imputing, the peaks in the training cohort ($n = 198$) were used for feature selection by three machine learning methods, including least absolute shrinkage and selection operator (LASSO), partial least-squares-discriminant analysis (PLS-

DA), and recursive feature elimination with cross-validation (RFECV). The 20 peaks with the highest repetition frequency in LASSO, 20 peaks with the highest variable importance in projection (VIP) in PLS-DA, and 10 peaks with automatic tuning of all the features selected by cross-validation accuracy in RFECV were chosen (Figure 2b–d). After further empirical checking, 25 peaks were identified as the distinctive features between COVID-19 patients and control participants. The intensities of all the feature peaks in the training cohort are shown in Figure 2e and Figure S2 and Table S2. The intensities of these peaks were significantly different between the COVID-19 and control groups.

Then, the classification models for the detection of COVID-19 were constructed with the 25 feature peaks in the training cohort by cross validation (9:1, 10 times) using eight machine learning methods, including logistic regression (LR), support vector machine (SVM), random forest (RF), naive Bayes (NB), gradient boosting decision tree (GBDT), K-nearest neighbor (KNN), decision tree (DT), and adaptive boosting (Adaboost). The receiver operating characteristic (ROC) curves of the different machine learning models are shown in Figure 2f. The area under the curve (AUC) evaluates the performance of a classifier, and a higher AUC means a better classification.³⁹ All of the models got AUC above 0.99. For LR, SVM, RF, GBDT, DT, and Adaboost models, the AUC was 1.

Identification of COVID-19 Patients by MALDI-TOF Profiling of Serum Using Machine Learning-Based Classification. The classification efficiency of the 25 feature peaks was then validated in a test cohort ($n = 100$) that was independent from the samples included in the training cohort.

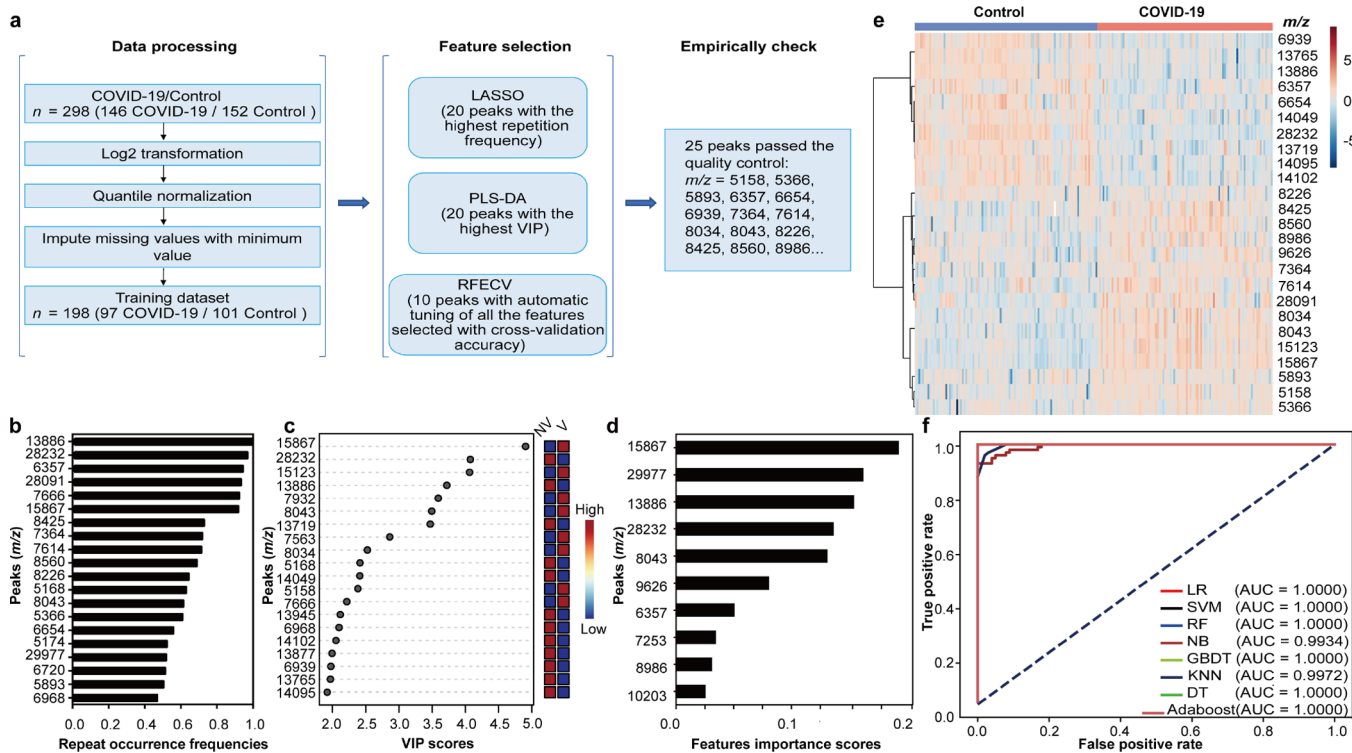


Figure 2. Selection of 25 feature peaks for COVID-19 detection. (a) General scheme of the data processing and feature selection workflow. (b) Top 20 features prioritized by LASSO analysis ranked by the decrease in repetition frequency. (c) Top 20 features prioritized by PLS-DA ranked by the decrease in VIP values. NV: non-COVID-19. V: COVID-19. (d) Top 10 features prioritized by RFECV ranked by the decrease in feature importance scores. (e) Heatmap of the selected 25 features. (f) ROC curves of eight different machine learning models in the training cohort by cross validation.

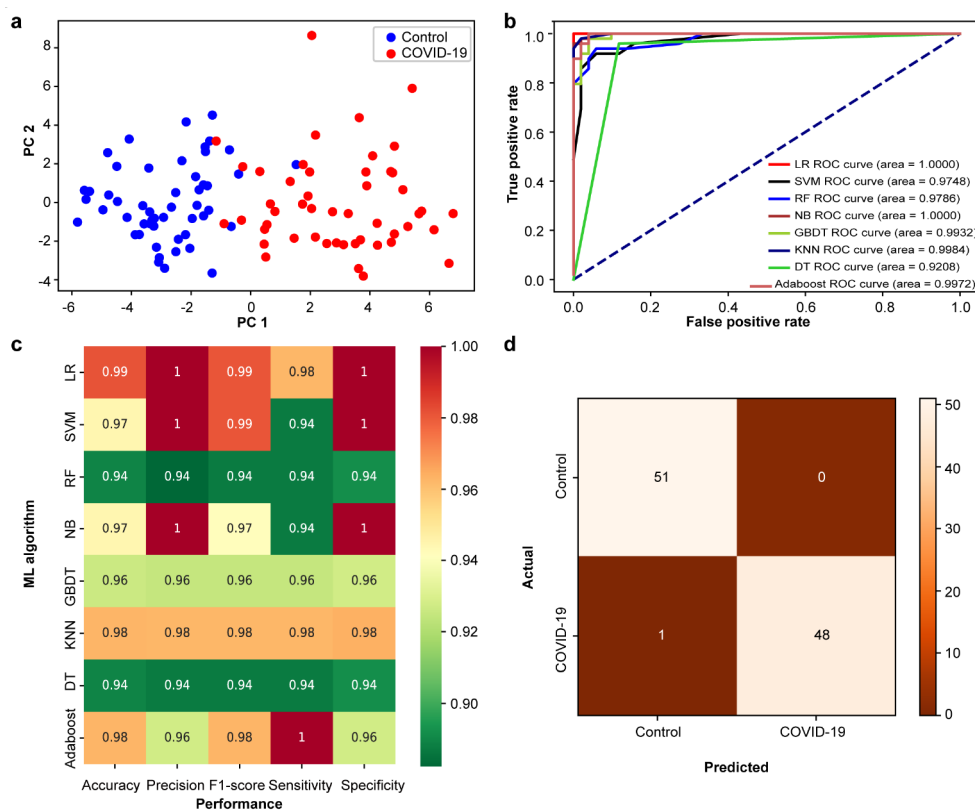


Figure 3. Identification of COVID-19 patients using machine learning-based classification modes in the test cohort. (a) PCA analysis using the 25 features. (b) ROC curves by eight machine learning methods. (c) Summary of the accuracy, precision, F1-score, sensitivity, and specificity obtained for each machine learning method. (d) Confusion matrix of the classification results by the LR mode.

The intensities of the feature peaks in the test cohort are shown in Table S3. The PCA analysis based on the 25 features demonstrated that the COVID-19 patients and control cases in the test group could be well separated (Figure 3a). All of the eight models from the eight machine learning methods established by using the training cohort obtained AUC above 0.92 in the test cohort (Figure 3b). For LR, SVM, and NB models, the AUC was 1. The sensitivity, specificity, accuracy, precision, and F1-score (the definitions of these terms listed in Text S1) of the eight models are shown in Figure 3c. With the best classification performance (AUC = 1, sensitivity = 98%, specificity = 100%, accuracy = 99%, precision = 100%, F1-score = 99%), the LR model is recommended for future applications in the detection of COVID-19. The confusion matrix of the LR model in the test cohort is shown in Figure 3d, and only one sample was misclassified among 100 samples. We also tested the feasibility of classifying the 146 COVID-19 patients with different degrees of severity based on the MALDI-TOF spectra. There was no obvious difference between the different groups by considering all the aligned peaks or the 25 feature peaks.

The symptoms of COVID-19 patients usually are non-specific. Guan et al. reported that only 44% of COVID-19 patients from China had a fever when they entered the hospital.¹⁴ At the same time, other symptoms such as cough, fatigue, sputum production, and shortness of breath could also be associated with other respiratory infections. It is very important to screen COVID-19 patients from non-COVID-19 individuals, especially with the similar symptoms. In this study, about half of the control participants were non-COVID-19

patients with similar clinical symptoms ($n = 73$), together with 33 tuberculosis patients and 46 healthy individuals. The sampling time of COVID-19 patients' sera analyzed in this study ranged from 3 to 28 days from the onset of symptoms, which covered a relatively long period of disease progression. The long disease progression coverage of the samples demonstrated the great potential of the method in the screening of COVID-19 patients.

We kept the most simplified experimental procedure, using the typical MALDI-TOF method and with only minimal sample preparation steps, in order to develop a high-throughput method that can be mastered by nonexperts. Because the organic matrix used in typical MALDI-TOF analysis can generate strong background signals in low mass regions, metabolites were not considered for MS analysis. Without proteolysis, large proteins cannot be efficiently ionized and detected. Hence, the mass range of the analysis focused on 5000 to 30,000 m/z , corresponding to serum peptidome and small proteins. Compared with the commonly used virus nucleic acids detection methods based on PCR, the MALDI-TOF MS-based serum peptidome profiling does not need an extra clean testing environment and has a low risk of sample cross contamination. Only 5 μL of serum sample was used per analysis. The serum samples could be analyzed immediately after sterilization. It took less than 1 min to analyze a sample by the MALDI-TOF MS. Without expensive consumables, the cost per test by MALDI-TOF MS is much lower compared to PCR or immunoassay, i.e., less than 1 US dollar per sample. Nachtigall et al. have recently reported the detection of COVID-19 using MALDI-TOF MS analysis of nasal swab

samples and found seven feature peaks to distinguish COVID-19 patients from the control participants.⁴⁰ Compared to nasal swabs, the serum sample is more static and less sensitive to external turbulence for more reliable diagnosis. Furthermore, serum collection also reduced the exposure risk of samplers.

Annotation of Feature Peaks. To identify the feature peaks, the sera from four COVID-19 patients (VSamples 67, 81, 112, 140) and three control participants (Samples 36, 41, 145) were selected for proteomic analysis. A total of 753 proteins were identified from the sera samples by LC-MS/MS and then matched against the 25 feature peaks under the criteria detailed in Text S1. Fifteen out of the 25 feature peaks were identified as intact proteins or protein fragments (Table S4). As shown in Figure S3, the top Gene Ontology (GO) processes enriched by Metascape⁴¹ involving the 15 identified features include amyloid fiber formation, neutrophil degranulation, infection with *Mycobacterium tuberculosis*, humoral immune response, receptor-mediated endocytosis, acute inflammatory response, and regulation of MAP kinase activity. In the previous proteomic and metabolomic study of COVID-19 patients' sera by Shen et al., differentially expressed proteins in COVID-19 patients enriched in the function of neutrophil degranulation and acute inflammatory response were also reported.⁴² Seven peaks were identified as proteins or peptides with the function of humoral immune response. Among the proteins, platelet factor 4 and complement factor I were investigated by Shen et al.⁴² and Shu et al.,⁴³ respectively, among COVID-19 patients with different severities. Gordon et al. have identified the gene RAB7A as a host factor with functional relevance in SARS-CoV-2 infections.⁴⁴ In our result, Ras-related protein Rab-7a encoded by RAB7A was identified from one of the feature peaks. These results demonstrate the rationality of the feature peaks in the identification of COVID-19.

CONCLUSION

In summary, our data demonstrated that MALDI-TOF-based serum profiling is a rapid and accurate method for the detection of COVID-19. It has great potential for screening, routine surveillance, and diagnostic applications in large populations, which is an important part for the pandemic control. Future studies would include the asymptomatic infected COVID-19 patients, which was absent in the current work due to the limitation of sample availability. The performance of the LR model used in this study would also be validated in other cohorts.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.0c04590>.

Supplementary methods, representative MALDI-TOF mass spectra of samples from COVID-19 patients and control participants, demographics of COVID-19 patients and controls, relative intensity comparison of the 25 feature peaks between controls and COVID-19 patients in the training cohort and test cohort, and identification results of the feature peaks (PDF)

Data set S1: Clinical information on cases enrolled in this study (XLSX)

Data set S2: Processed MALDI-MS intensity matrix (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Liang Qiao – Department of Chemistry, Fudan University, Shanghai 200438, China; orcid.org/0000-0002-6233-8459; Email: liang_qiao@fudan.edu.cn

Wei Sun – State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China; Email: sunway21cn@126.com

Pu Liao – Department of Clinical Laboratory, Chongqing General Hospital, Chongqing 400014, China; Email: liaopu@sina.com

Authors

Ling Yan – Department of Clinical Laboratory, Chongqing General Hospital, Chongqing 400014, China

Jia Yi – Department of Chemistry, Fudan University, Shanghai 200438, China

Changwu Huang – Department of Clinical Laboratory, Renji Hospital, University of Chinese Academy of Sciences, Chongqing, China

Jian Zhang – State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China

Shuhui Fu – Bioyong Technologic, Inc., Beijing 100176, China

Zhijie Li – Department of Clinical Laboratory, Chongqing General Hospital, Chongqing 400014, China

Qian Lyu – Bioyong Technologic, Inc., Beijing 100176, China

Yuan Xu – Department of Clinical Laboratory, Chongqing General Hospital, Chongqing 400014, China

Kun Wang – Department of Clinical Laboratory, Chongqing General Hospital, Chongqing 400014, China

Huan Yang – Department of Clinical Laboratory, Chongqing General Hospital, Chongqing 400014, China

Qingwei Ma – Bioyong Technologic, Inc., Beijing 100176, China

Xiaoping Cui – Department of Clinical Laboratory, Sanxia Hospital Affiliated to Chongqing University, Chongqing 404000, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.0c04590>

Author Contributions

Sample collection was performed by C.H., Z.L., K.W., and X.C. under the supervision of P.L. MALDI-TOF experiments were carried out by L.Y. and Z.L. with the assistance of Q.L. and Q.M. Y.X. and H.Y. performed the sample preparation for proteomic analysis. Data analyses were performed by J.Y., J.Z., and S.F. under the supervision of L.Q. and W.S. The paper was written with contributions from all authors.

Author Contributions

[∇]L. Yan, J. Yi, C. Huang, and J. Zhang contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was funded by the National Megaprojects for Key Infectious Diseases (2018ZX10732202-003), the Ministry of Science and Technology (MOST) of the People's Republic of China (SQ2020YFF0426500, 2017YFC0906702), and the

Beijing Municipal Science & Technology Commission (Z181100001618015).

REFERENCES

- (1) Zhou, P.; et al. *Nature* **2020**, 579 (7798), 270–273.
- (2) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. *N. Engl. J. Med.* **2020**, 382 (8), 727–733.
- (3) World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (accessed February 2020).
- (4) 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. *ACS Nano* **2020**, 14 (4), 3822–3835.
- (5) World Health Organization. *Laboratory Testing for Coronavirus Disease 2019 (COVID-19) in Suspected Human Cases: Interim Guidance, 2 March 2020*; World Health Organization: Geneva, 2020.
- (6) Ai, T.; Yang, Z.; Hou, H.; Zhan, C.; Chen, C.; Lv, W.; Tao, Q.; Sun, Z.; Xia, L. *Radiology* **2020**, 296 (2), E32–E40.
- (7) Winichakoon, P.; Chaiwarit, R.; Liwsrisakun, C.; Salee, P.; Goonna, A.; Limsukon, A.; Kaewpoowat, Q. *J. Clin. Microbiol.* **2020**, 58 (5), e00297-20.
- (8) Zhang, W.; Du, R. H.; Li, B.; Zheng, X. S.; Yang, X. L.; Hu, B.; Wang, Y. Y.; Xiao, G. F.; Yan, B.; Shi, Z. L.; Zhou, P. *Emerging Microbes Infect.* **2020**, 9 (1), 386–389.
- (9) Ackerman, C. M.; Myhrvold, C.; Thakku, S. G.; Freije, C. A.; Metsky, H. C.; Yang, D. K.; Ye, S. H.; Boehm, C. K.; Kosoko-Thoroddsen, T.-S. F.; Kehe, J.; Nguyen, T. G.; Carter, A.; Kulesa, A.; Barnes, J. R.; Dugan, V. G.; Hung, D. T.; Blainey, P. C.; Sabeti, P. C. *Nature* **2020**, 582 (7811), 277–282.
- (10) Yu, L.; Wu, S.; Hao, X.; Dong, X.; Mao, L.; Pelechano, V.; Chen, W.-H.; Yin, X. *Clin. Chem.* **2020**, 66 (7), 975–977.
- (11) Ihling, C.; Taenzler, D.; Hagemann, S.; Kehlen, A.; Huettelmaier, S.; Arlt, C.; Sinz, A. *J. Proteome Res.* **2020**, 19 (11), 4389–4392.
- (12) Nikolaev, E. N.; Indeykina, M. I.; Brzhozovskiy, A. G.; Bugrova, A. E.; Kononikhin, A. S.; Starodubtseva, N. L.; Petrotchenko, E. V.; Kovalev, G. I.; Borchers, C. H.; Sukhikh, G. T. *J. Proteome Res.* **2020**, 19 (11), 4393–4397.
- (13) Long, Q. X.; et al. *Nat. Med.* **2020**, 26, 845–848.
- (14) Guan, W.; et al. *N. Engl. J. Med.* **2020**, 382 (18), 1708–1720.
- (15) Fang, Y.; Zhang, H.; Xie, J.; Lin, M.; Ying, L.; Pang, P.; Ji, W. *Radiology* **2020**, 296 (2), E115–E117.
- (16) Xie, X.; Zhong, Z.; Zhao, W.; Zheng, C.; Wang, F.; Liu, J. *Radiology* **2020**, 296 (2), E41–E45.
- (17) Yang, Y.; et al. *J. Allergy Clin. Immunol.* **2020**, 146 (1), 119–127.
- (18) Tirumalai, R. S.; Chan, K. C.; Prieto, D. A.; Issaq, H. J.; Conrads, T. P.; Veenstra, T. D. *Mol. Cell. Proteomics* **2003**, 2 (10), 1096–1103.
- (19) Villanueva, J.; Shaffer, D. R.; Philip, J.; Chaparro, C. A.; Erdjument-Bromage, H.; Olshen, A. B.; Fleisher, M.; Lilja, H.; Brogi, E.; Boyd, J.; Sanchez-Carbayo, M.; Holland, E. C.; Cordon-Cardo, C.; Scher, H. I.; Tempst, P. *J. Clin. Invest.* **2005**, 116 (1), 271–284.
- (20) Claydon, M. A.; Davey, S. N.; Edwards-Jones, V.; Gordon, D. B. *Nat. Biotechnol.* **1996**, 14 (11), 1584–1586.
- (21) Butts, C. A. *Lancet Oncol.* **2014**, 15 (7), 671–672.
- (22) Nakamura, A.; Kaneko, N.; Villemagne, V. L.; Kato, T.; Doecke, J.; Dore, V.; Fowler, C.; Li, Q.-X.; Martins, R.; Rowe, C.; Tomita, T.; Matsuzaki, K.; Ishii, K.; Ishii, K.; Arahata, Y.; Iwamoto, S.; Ito, K.; Tanaka, K.; Masters, C. L.; Yanagisawa, K. *Nature* **2018**, 554 (7691), 249–254.
- (23) Liu, C.; Zhao, Z.; Fan, J.; Lyon, C. J.; Wu, H. J.; Nedelkov, D.; Zelazny, A. M.; Olivier, K. N.; Cazares, L. H.; Holland, S. M.; Graviss, E. A.; Hu, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, 114 (15), 3969–3974.
- (24) Gibb, S.; Strimmer, K. *Bioinformatics* **2012**, 28 (17), 2270–2271.
- (25) Bolstad, B. M.; Irizarry, R. A.; Astrand, M.; Speed, T. P. *Bioinformatics* **2003**, 19 (2), 185–193.
- (26) Chong, J.; Wishart, D. S.; Xia, J. *Curr. Protoc. Bioinformatics* **2019**, 68 (1), e86.
- (27) Tibshirani, R. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **2011**, 73, 273–282.
- (28) Friedman, J.; Hastie, T.; Tibshirani, R. *J. Stat. Softw.* **2010**, 33 (1), 1–22.
- (29) Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. *MLear* **2002**, 46 (1–3), 389–422.
- (30) Fan, R.-E.; Chang, K.-W.; Hsieh, C.-J.; Wang, X.-R.; Lin, C.-J. *J. Mach. Learn. Res.* **2008**, 9, 1871–1874.
- (31) Platt, J. C. *Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods*; MIT Press, 1999.
- (32) Chan, T. F.; Golub, G. H.; LeVeque, R. J. Updating Formulae and a Pairwise Algorithm for Computing Sample Variances. In *COMPSTAT 1982 5th Symposium*, Toulouse, 1982; pp 30–41.
- (33) Breiman, L.; Friedman, J.; Stone, C. G.; Olshen, R. A. *Classification and Regression Trees*; CRC, 1984.
- (34) Breiman, L. *MLear* **2001**, 45 (1), 5–32.
- (35) Friedman, J. H. *Ann. Stat.* **2001**, 29 (5), 1189–1232.
- (36) Friedman, J. *Stochastic Gradient Boosting*; Department of Statistics, Stanford University, 1999.
- (37) Zhu, J.; Zou, H.; Rosset, S.; Hastie, T. *Stat. Interface* **2009**, 2 (3), 349–360.
- (38) Wang, H. *Nearest Neighbours without k: A Classification Formalism Based on Probability*; Faculty of Informatics, University of Ulster, 2002.
- (39) Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.-C.; Mueller, M. *BMC Bioinf.* **2011**, 12, 77.
- (40) Nachtigall, F. M.; Pereira, A.; Trofymchuk, O. S.; Santos, L. S. *Nat. Biotechnol.* **2020**, 38, 1168–1173.
- (41) Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. *Nat. Commun.* **2019**, 10, 1523.
- (42) Shen, B.; et al. *Cell* **2020**, 182 (1), 59–72.
- (43) Shu, T.; et al. *Immunity* **2020**, 53 (5), 1108–1122.
- (44) Gordon, D. E.; et al. *Science* **2020**, 370 (6521), eabe9403.