Journal of proteome research

pubs.acs.org/jpr Article

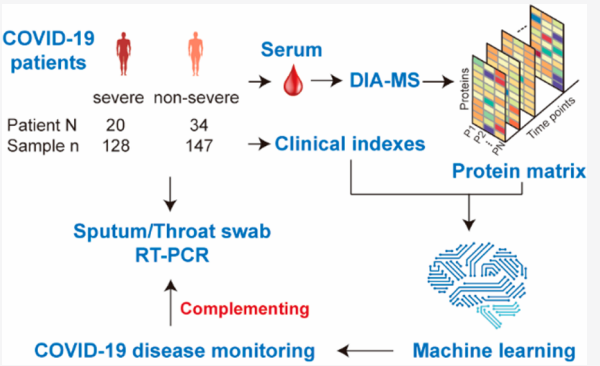
Potential Use of Serum Proteomics for Monitoring COVID-19 Progression to Complement RT-PCR Detection

Ying Zhang,[#] Xue Cai,[#] Weigang Ge,[#] Donglian Wang,[#] Guangjun Zhu,[#] Liujia Qian,[#] Nan Xiang, Liang Yue, Shuang Liang, Fangfei Zhang, Jing Wang, Kai Zhou, Yufen Zheng, Minjie Lin, Tong Sun, Ruyue Lu, Chao Zhang, Luang Xu, Yaoting Sun, Xiaoxu Zhou, Jing Yu, Mengge Lyu, Bo Shen, Hongguo Zhu,* Jiaqin Xu,* Yi Zhu,* and Tiannan Guo*

Cite This: https://doi.org/10.1021/acs.jproteome.1c00525 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: RT-PCR is the primary method to diagnose COVID-19 and is also used to monitor the disease course. This approach, however, suffers from false negatives due to RNA instability and poses a high risk to medical practitioners. Here, we investigated the potential of using serum proteomics to predict viral nucleic acid positivity during COVID-19. We analyzed the proteome of 275 inactivated serum samples from 54 out of 144 COVID-19 patients and shortlisted 42 regulated proteins in the severe group and 12 in the non-severe group. Using these regulated proteins and several key clinical indexes, including days after symptoms onset, platelet counts, and magnesium, we developed two machine learning models to predict nucleic acid positivity, with an AUC of 0.94 in severe cases and 0.89 in non-severe cases, respectively. Our data suggest the potential of using a serum protein-based machine learning model to monitor COVID-19 progression, thus complementing swab RT-PCR tests. More efforts are required to promote this approach into clinical practice since mass spectrometry-based protein measurement is not currently widely accessible in clinic.



KEYWORDS: *COVID-19, serum, proteomics, disease course monitoring, machine learning*

INTRODUCTION

Since December 2019, the COVID-19 outbreak, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2),[1] has developed into a worldwide pandemic and shown a high risk of person-to-person transmission.[2,3] The discharge criteria for COVID-19 patients, as summarized by the COVID-19 Diagnosis and Treatment Protocol (trial version 5),[4] includes: "nucleic acid tests negative twice consecutively on respiratory tract samples such as sputum and nasopharyngeal swabs (sampling interval being at least 24 h)".

The pharyngeal or nasopharyngeal swabs of patients are thus collected by clinicians for RT-PCR detection, as often as once a day in many hospitals,[5,6] which not only results in unnecessary consumption of medical resources but also increases the risk of infection for medical practitioners. The respiratory tract samples of COVID-19 patients are more likely to have live SARS-CoV-2 viruses than blood and feces samples.[7] The direct exposure to swab samples with a high viral load during sample collection thus poses a significant risk to the practitioners involved in the sampling and the sample analysis.[4,8] In addition, due to the RNA instability, false-negative results from RT-PCR are frequent.[9,10] The storage and transfer of these samples also require the highest level of biosafety considerations, further increasing the burden on medical resources.

Blood tests of multiple biochemical indexes are routinely performed for COVID-19 patients to offer a potential alternative for monitoring the disease progression. Recently, we reported a machine learning model for the classification of the severity of COVID-19 patients using 11 clinical features, six of which can be measured from blood samples,[11] and achieved an accuracy of 98% in the training set, 86% in a test set, and 80% in an independent test set. Another study characterized the molecular changes in heat-inactivated sera between severe and non-severe COVID-19 patients and built a machine learning model based on proteomics and metabolomics signatures to distinguish severe from non-severe COVID-19 patients.[12] These data showed that the heat inactivation of serum samples did not affect

Received: June 26, 2021

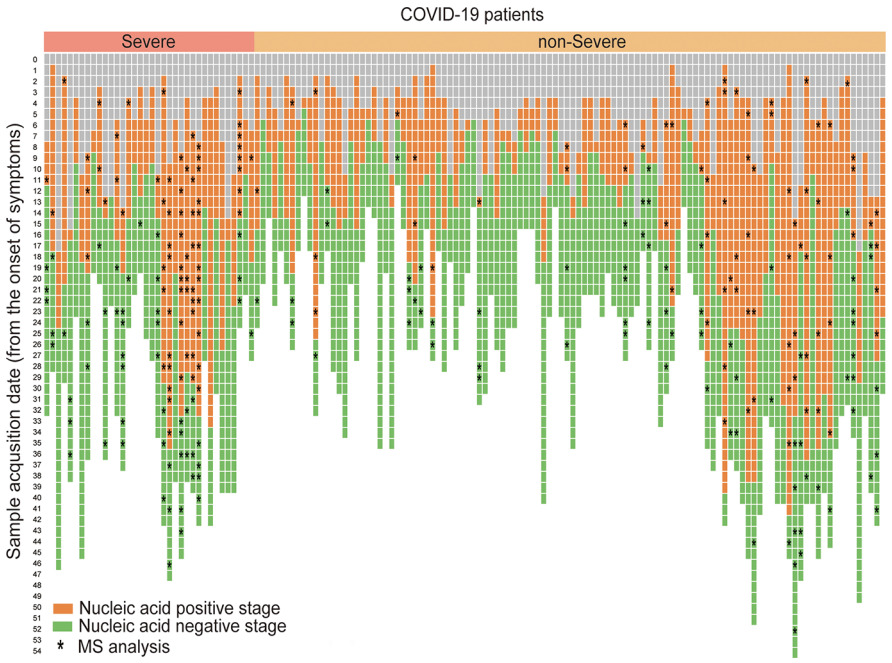


Figure 1. Summary of the serum sample collection from COVID-19 patients. The days are numbered from the onset of the symptoms. The gray squares represent the days before admission, the orange ones represent the days on which nucleic acid test was positive, and the green ones represent the days on which nucleic acid test was negative. A total of 275 serum samples were collected from 54 COVID-19 patients at different time points for MS analysis as indicated with ∗.

Table 1. Clinical Characteristics of COVID-19 Patients and Controls

variables	COVID-19			non-COVID-19 (n = 24)	healthy control (n = 21)
	total (n = 144)	non-severe (n = 108)	severe (n = 36)		
Sex- No. (%)					
male	77 (53)	57 (53)	20 (56)	16 (67)	14 (67)
female	67 (47)	51 (47)	16 (44)	8 (33)	7 (33)
Age- Yr.					
mean ± SD	47.7 ± 14.5	45.0 ± 14.2	55.7 ± 12.5	49.3 ± 14.3	45.2 ± 8.0
median (IQR)	47.0 (38.0−56.0)	44.5 (37.0−54.0)	55.0 (47.8−65.0)	54.0 (36.5−61.0)	46.0 (38.0−51.5)
range	4.0−86.0	4.0−86.0	33.0−79.0	23.0−67.0	28.0−57.0
Time from Onset to Admission, Days					
mean ± SD	7.0 ± 4.1	6.7 ± 3.8	7.9 ± 4.9		
median (IQR)	6.0 (4.0−10.0)	6.0 (4.0−9.0)	7.5 (4.0−11.0)		
range	1.0−24.0	1.0−18.0	1.0−24.0		
Time from Admission to Severe, Days					
mean ± SD			2.6 ± 1.5		
median (IQR)			2.0 (1.0−3.8)		
range			0.0−7.0		
Time from Admission to Discharge, Days					
mean ± SD	21.6 ± 9.4	20.5 ± 9.7	24.7 ± 7.8		
median (IQR)	21.5 (13.0−28.0)	20.0 (13.0−27.0)	23.0 (19.3−31.8)		
range	6.0−44.0	6.0−44.0	9.0−40.0		
Symptoms- No. (%)					
with fever	104 (72.2)	70 (64.8)	34.0 (94.4)		
without fever	40 (27.8)	38 (35.2)	2 (5.6)		
Comorbidity- No. (%)					
with comorbidity	59 (41.0)	41 (37.9)	18 (50.0)		
without comorbidity	85 (59.0)	67 (62.1	18 (50.0)		
Chest CT- No. (%)					
abnormal chest radiographs	141 (97.9)	105 (97.2)	36 (100.0)		
total sample- no.	631	380	251	24	21
sample with MS analysis- no.	275	147	128	24	21

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

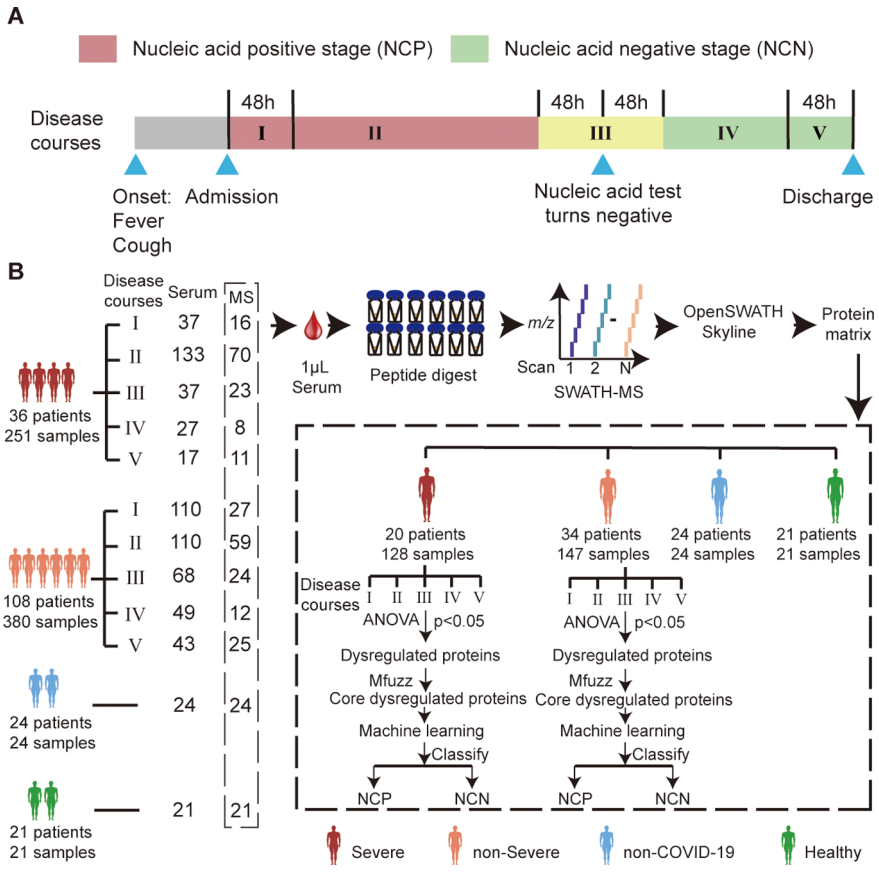


Figure 2. Study design. (A) Five stages of the COVID-19 course. Different colors represent different stages. (B) Workflow of the SWATH-MS and the data analysis. Study population: 36 severe and 108 non-severe COVID-19 patients, 24 non-COVID patients, and 21 healthy individuals. A total of 320 serum samples were analyzed by SWATH-MS. Dysregulated serum proteins were analyzed by ANOVA and Mfuzz in the five stages of the severe and non-severe COVID-19 cases, respectively. On the basis of the resulting dysregulated proteins, two machine learning models were built to identify NCP and NCN.

proteome profiling. Together, these studies suggested that proteomics profiling coupled with machine learning could effectively facilitate COVID-19 patient stratification.

In addition, dynamic monitoring of some clinical tests, such as antibody titers, expression of serum inflammatory markers,[13] X-ray,[14] and CT scan,[15] are empirically used to monitor the COVID-19 disease progression. This study investigated the potential of using blood protein measurement and machine learning to monitor the course of COVID-19.

MATERIALS AND METHODS

Patients and Samples

In this study, the patient cohort was from the Taizhou Public Health Medical Center (TPHMC) of Taizhou Hospital between January 17 and March 10, 2020. During this period, 144 individuals diagnosed with COVID-19 based on the Diagnostic and Treatment Criteria (Trial version 5)[4] were involved (Figure 1, Table 1). These 144 COVID-19 patients were classified in two categories: non-severe (n = 108) and severe (n = 36) cases.[4] Patients were diagnosed as severe cases if they met any of the following conditions: (i) respiratory distress, with a respiratory rate ≥ 30/min; (ii) oxygen saturation ≤ 93% in resting-state; (iii) partial arterial oxygen pressure (PaO_2)/oxygen absorption concentration (FiO_2) ≤ 300 mmHg. The study also included 24 non-COVID-19 patients with flu-like symptoms and 21 healthy individuals as a control group. This study was approved by the Ethical/Institutional Review Boards of both Taizhou Hospital and Westlake University. Informed contents from patients were waived by the boards.

COVID-19 Disease Stages

In this study, the disease course of 144 COVID-19 patients was split into five stages based on their clinical phenotypes (Figure 2A). Stage 1 refers to the first 48 h after admission, while stage 3 covers 96 h, the 48 before and the 48 after the virus test turned from positive to negative. Stage 5 refers to the 48 h before discharge. Stage 2 refers to the period between stages 1 and 3, and stage 4 between stages 3 and 5. As shown in Figure 2A, stages 1 and 2 belong to the nucleic acid positive (NCP) stage (highlighted in pink), while stages 4 and 5 belong to the nucleic acid negative (NCN) one (highlighted in green).

Peptide Preparation from Serum Samples

Peptides were extracted from the serum samples as previously described.[12] Briefly, after inactivation and sterilization at 56 °C for 30 min, the serum samples were lysed using a lysis buffer (8 M urea in 100 mM ammonium bicarbonate, ABB) at 32 °C for 30 min. The serum proteins were then reduced and alkylated using 10 mM tris (2-carboxyethyl) phosphine (TCEP) and 40 mM iodoacetamide (IAA), respectively. Before the enzymatic digestion, 200 μL of 100 mM ABB was added to the samples to dilute the urea. The protein extracts were then digested with a two-step tryptic digestion, an enzyme-to-substrate ratio of 1:20 (final ratio 1:10), and each step at 32 °C for 60 min. The

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

digestion was then stopped by adjusting the pH to 2−3 using 1% trifluoroacetic acid (TFA). Peptides were next cleaned with C18 SOLAu columns (Thermo Fisher Scientific, San Jose, USA) before MS analysis.

Spectral Library for DIA

The nanoflow DIONEX UltiMate 3000 RSLCnano System (Thermo Fisher Scientific, San Jose, USA) with an XBridge Peptide BEH C18 column (300 Å, 5 mm × 4.6 mm × 250 mm) (Waters, Milford, MA, USA) was used for high-pH fractionation. A complex peptide sample was fractionated into 90 aliquots using a 5−35% acetonitrile (ACN) gradient in 10 mM ammonia (pH = 10.0) at a flow rate of 1 mL/min. The resulting 90 aliquots were next combined into 45 fractions. The peptides were dried and redissolved in 2% ACN/0.1% FA, and each of the 45 peptides fractions was injected into the Eksigent NanoLC 400 System (Eksigent, Dublin, CA, USA) coupled with TripleTOF 5600 and 6600 system (SCIEX, CA, USA) for information-dependent acquisition (IDA) MS analysis. The online HPLC was performed at a flow rate of 5 μL/min with a 90 min gradient of 5%−32% buffer B (buffer A: 2% ACN, 0.1% formic acid; buffer B: 98% ACN, 0.1% formic acid). In the IDA method, the accumulation time for the MS1 and MS2 scans was set to 250 and 50 ms, respectively. The MS/MS scans were performed for the top 40 precursors, resulting in a total cycling time of 2.3 s. All 45 fractions were acquired in a TripleTOF 6600 MS. Then 36 out of 45 fractions were further combined into 12 and analyzed in a TripleTOF 5600 MS. In total, 57 IDA wiff files were analyzed using Maxquant,[16,17] MSFragger,[18] pFind,[19,20] and merged to a serum spectral library containing 10 001 peptides and 2592 protein groups.[21] We next built a subset library containing 3474 peptides precursors and 536 protein groups for SWATH-MS data analysis, as previously described.[22]

SWATH-MS Analysis

The peptides samples were injected into an Eksigent NanoLC 400 System (Eksigent, Dublin, CA, USA), coupled with a TripleTOF 5600 system (SCIEX, CA, USA) for the SWATH-MS analysis. For each sample, 500 ng of peptides was separated along an analytical column (3 μm, ChromXP C18CL, 120 Å, 150 × 0.3 mm) with a 20 min LC gradient of 5−30% buffer B, at a flow rate of 5 μL/min. A 55 variable Q1 isolation window scheme was set for the SWATH method used in this study (Table S1A). The accumulation time was set to 100 ms for the MS1 scan and 31 ms for the MS/MS scan. The duty time per cycle was 1.8 s.

The SWATH wiff files were converted into mzXML format using msConvert[23] and analyzed using OpenSWATH (version 2.4)[24] against the serum spectral library established above. The retention time extraction window was set to 120 s, and the m/z extraction for MS1 and MS2 was performed at m/z tolerances of 20 and 50 ppm, respectively. The retention time was then calibrated using common internal retention time (CiRT) standards peptides (Table S1B).[25] The m/z extraction for CiRT peptides was performed at an m/z tolerance of 50 ppm. The peptides precursors were identified by OpenSWATH and pyprophet with FDR < 0.01. The C-reactive protein (CRP) data were manually analyzed with Skyline.[26] The retention time was predicted by the above-mentioned CiRT peptides, and the isolation time window was set to 2 min. The mass analyzer for MS1 and MS/MS was set to "TOF", with a resolution power of 30 000.

Clinical Assays

The serum concentrations of serum amyloid A (SAA) (mg/L) and CRP (mg/L) were measured with the immunoturbidimetric method (the kit for SAA protein assay from Ningbo Purebio Biotechnology co. Ltd., and the kit For CRP Assay from Beckman Coulter). The serum concentrations of triglyceride (TRIG) (mmol/L) and low-density lipoprotein (LDL) (mmol/L) were assayed with the glycerophosphate oxidase peroxidase (GPO-PAP) and the creatine phosphate substrate method, respectively. The kits for TRIG and LDL assays were both from Beijing Leadman Biochemistry Co., Ltd. These assays were performed on an AU5821 analyzer (Beckman Coulter, California, USA). The mean platelet volume (MPV) (fL) was measured using K2 EDTA anticoagulant peripheral blood samples with the Coulter principle on a 2100D analyzer (Sysmex, Kobe, Japan).

Statistical Analysis and Machine Learning

The dysregulated serum proteins along with five disease stages (Figure 2A) were analyzed using ANOVA. The Pearson correlation coefficient (r) was calculated by the "cor" function in R (version 3.6.1) with the missing value options set to "pairwise.complete.obs". The random forest analysis was performed with the R package randomForest (version 4.6.14) as previously described[12] with some modifications. The key random forest parameters, including the cutoff values, were optimized for the decreasing mean accuracy, the cross-validation fold, and the number of trees. We selected the input protein features and the clinical features for machine learning based on the decreasing mean accuracy cutoff. Six-fold cross-validation was performed and repeated 50 times to optimize the model. A total of 600 trees were built. The minimal decreasing mean accuracy of protein and clinical features for severe models was set to 3 and for non-severe ones to 0. The mtry for severe and non-severe models was set to the square root of 8 and 23, respectively.

RESULTS AND DISCUSSION

Study Design

For the classification as nucleic acid positive (NCP) or nucleic acid negative (NCN) using machine learning models, we used the clinical data derived from 631 serum samples collected from the 144 patients across these five stages. We also analyzed the proteome of 275 samples, based on sample availability, using SWATH-MS. Additionally, we collected 45 samples from 24 flu-like non-COVID-19 patients and 21 healthy individuals, as controls. The detailed information on these patients and samples is provided in Table 1 and Table S2A. The average age of the COVID-19 patients was 47.7 ± 14.5 years, with 77 (53.5%) male and 67 (46.5%) female patients. Fever was their most common symptom (n = 104, 72.2%). The average interval from the onset of the symptoms to the hospital admission was 6 days (IQR 4−10). The median length of hospitalization was 21.5 days (IQR 13−28). A total of 141 (97.9%) patients had abnormal chest radiographs, and 59 (41.0%) had coexisting conditions.

Temporal Proteomics Profiling of 320 Serum Samples by SWATH-MS

To obtain the temporal proteomics profiling data, we analyzed 320 serum samples using microflow short-gradient single-shot SWATH-MS.[27] These samples included 128 from 20 severe patients, 147 from 34 non-severe patients, 24 from 24 non-

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

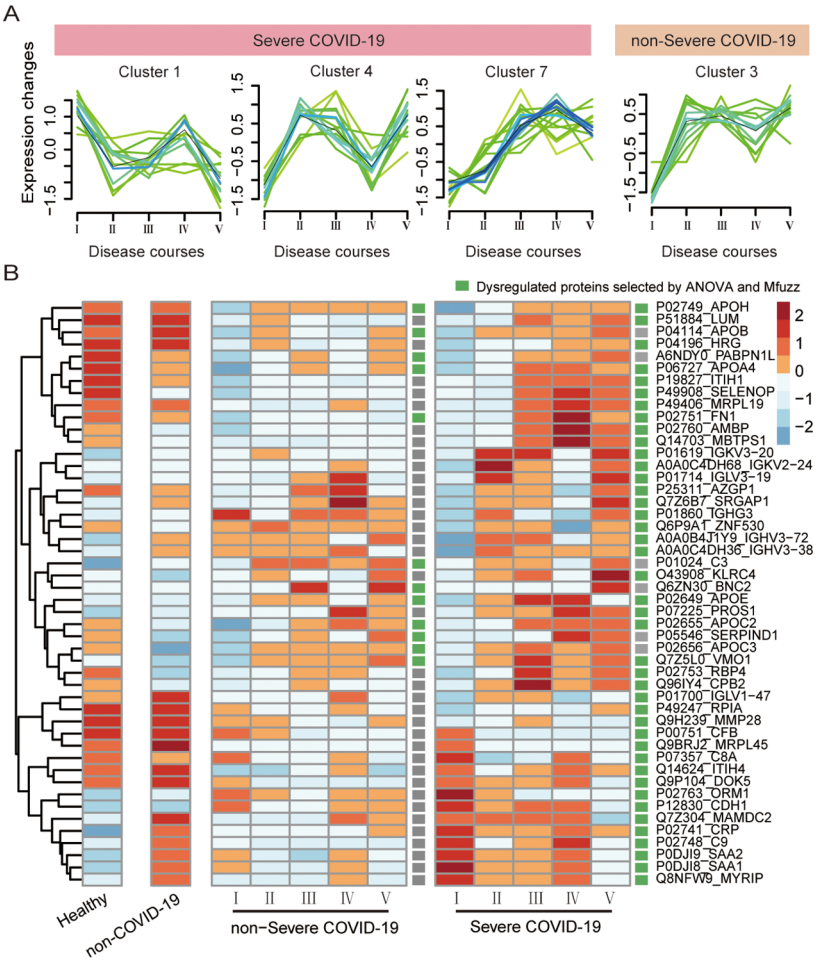


Figure 3. Dysregulated serum proteins dynamics in the course of COVID-19. (A) Three clusters of proteins identified with Mfuzz analysis showed trends of continuous changes in the five stages of severe COVID-19 course. One cluster showed an upward trend for the non-severe COVID-19 course. (B) Heatmap of 48 dysregulated proteins in the five stages of the COVID-19 course. Proteins with green boxes were selected as input features for the machine learning models using ANOVA and Mfuzz.

COVID-19 patients, and 21 from 21 healthy individuals (Figure 1, Figure 2B, Table S3A). A protein matrix including 337 serum proteins, with 21.5% missing values, was generated (Table S3B) using OpenSWATH (version 2.4), with an in-house serum spectral library containing 3474 peptide precursors and 536 protein groups using the SubLib strategy.[22]

During the SWATH data acquisition, we included seven pooled samples injected every 50 samples for the quality control (QC) of the MS. The Pearson coefficient of these QC samples was 0.87 (Figure S1). We also included 27 pairs of randomly selected biological replicates and 22 pairs of randomly selected technical replicates, once again for quality assessment. The median Pearson correlation coefficients of these biological and technical replicates were 0.87 and 0.90 (Figure S2), respectively, suggesting a relatively high degree of consistency and reproducibility.

ANOVA was used to identify the dysregulated proteins during the five stages of the entire COVID-19 course for the severe and the non-severe patients. We found that 120 and 31 proteins were significantly dysregulated in the severe and non-severe groups, respectively (p-value < 0.05). Mfuzz analysis[28] was then used to cluster the dysregulated proteins derived from the ANOVA analysis. A total of 42 and 12 proteins exhibited consistent down- or up-regulation, respectively, during the five disease stages (Figure 3A, Table S3C). The resulting 48 dysregulated and unique proteins were then run through the machine learning analysis described in the next section (Figure 3B, Table S3C). We also explored the perturbed pathways over the course of COVID-19 for the 48 dysregulated proteins using Ingenuity Pathway Analysis (IPA), and nine pathways were enriched (Figure S3). Among them, four pathways, including LXR/RXR activation, FXR/RXR activation, IL-12 signaling and production in macrophages, and production of nitric oxide and reactive oxygen species in macrophages, were related to host defense, which might influence the outcome of the disease. We also observed dysregulation of the complement system, which was reported to be associated with the severity of COVID-19.[12,29,30]

Additionally, we analyzed the correlation between the 48 dysregulated proteins (Table S3C) and 54 clinical indexes for COVID-19 patients. A total of 10 dysregulated proteins highly correlated with 16 clinical indexes (Table S2B) have been previously reported to be associated with COVID-19 severity (Figure S4).[11,31−36] This result further corroborates these proteins as potential biomarkers for the monitoring of the disease progression.

Classification of COVID-19 Stages with Machine Learning

We next built two machine learning models to investigate the possibility of distinguishing the NCP and NCN stages of severe and non-severe COVID-19 patients, respectively (Figure 4A).

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

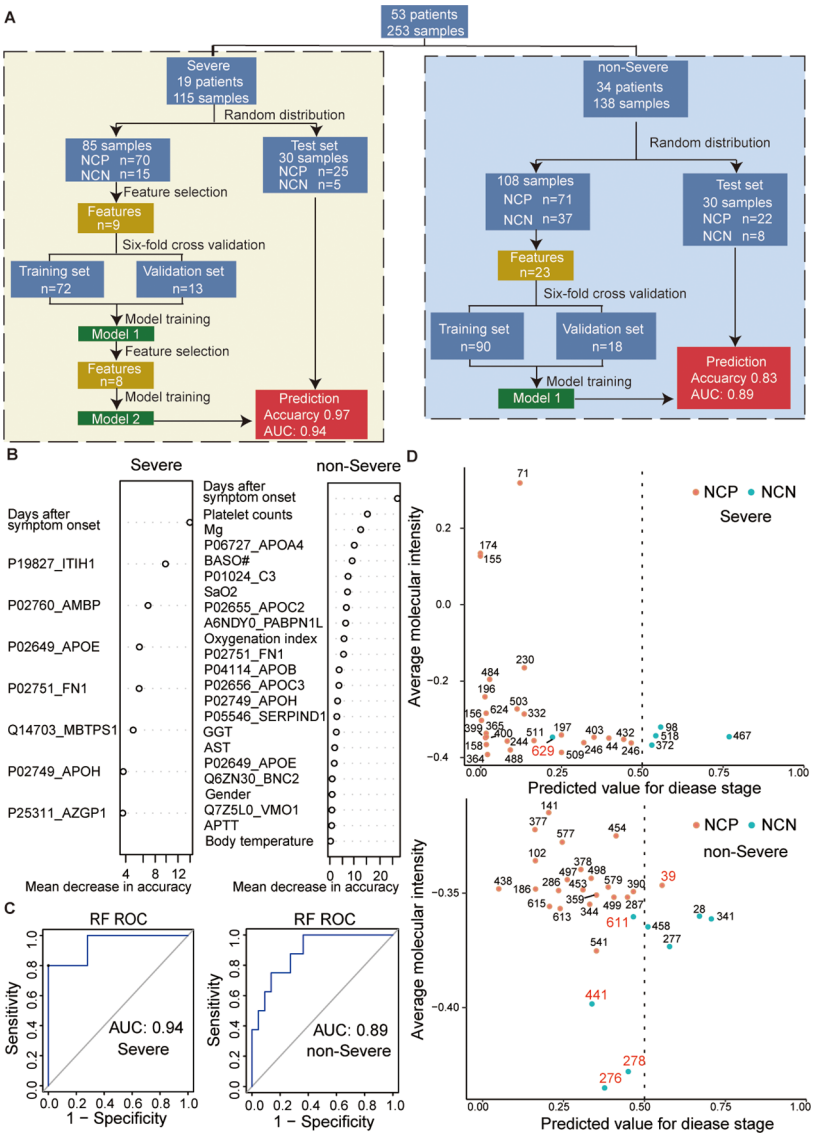


Figure 4. Machine learning models for predicting the stage of the severe and non-severe courses. (A) Workflow of two machine learning models built with proteomics quantitative data and clinical indexes. (B) Prioritization of 8 important variables in the severe model and 23 important variables in the non-severe model. (C) ROC plots for the two machine learning models of the severe (left) and the non-severe patients (right). (D) Performance of the severe and the non-severe models in the two test cohorts of 30 independent samples. The labeled numbers represent the sample ID.

Toward this goal, we used the dysregulated proteins previously identified in our proteomics analysis and other relevant clinical indexes. One patient only had one sample for stage 3 and thus was excluded from this analysis.

For the severe group, we built a random forest model based on the serum proteome data (n = 85) and a set of characteristic features including the 42 dysregulated proteins (Table S3C) and 11 clinical indexes selected in our previous study.[37] Specifically, the 11 clinical indexes were the oxygenation index, the basophil count (BASO#), aspartate aminotransferase (AST), gender, magnesium (Mg), gamma-glutamyl transpeptidase (GGT), the platelet count, the activated partial thromboplastin time (APTT), the oxygen saturation (SaO_2), body temperature, and the days after the symptoms' onset. After a 6-fold cross-validation using the training set and feature selection, we identified the eight features providing the highest accuracy for the severe cases (mean decrease accuracy > 0, which represents the loss of accuracy): interalpha-trypsin inhibitor heavy chain H1 (ITIH1), alpha-1-microglobulin (AMBP), apolipoprotein E (APOE), fibronectin (FN1), membrane-bound transcription factor site-1 protease (MBTPS1), apolipoprotein H (APOH), zinc-alpha-2-glycoprotein (AZGP1), and the days after the symptoms' onset (Figure 4B).

We thus tested this model using 30 independent samples and achieved an area under the curve (AUC) of 0.94 (Figure 4C). All samples were correctly classified, except one (CVDSBB629) belonging to the NCN stage (prediction accuracy = 0.97). This sample, derived from a 77-year old severe female patient with hypertension and diabetes, was incorrectly classified by our machine learning model to belong to the NCP stage (Figure 4D). During her hospitalization, this patient was diagnosed with renal insufficiency, type 1 respiratory failure, cardiac insufficiency, hypoproteinemia, and fungal infection. Multiple pathological alterations were also discovered including aortic wall calcification, cystic lesions of the pancreas, and small stones in the left kidney in this patient. These multiorgan dysfunctions and her treatment history may thus have compromised the model's prediction for this patient.

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

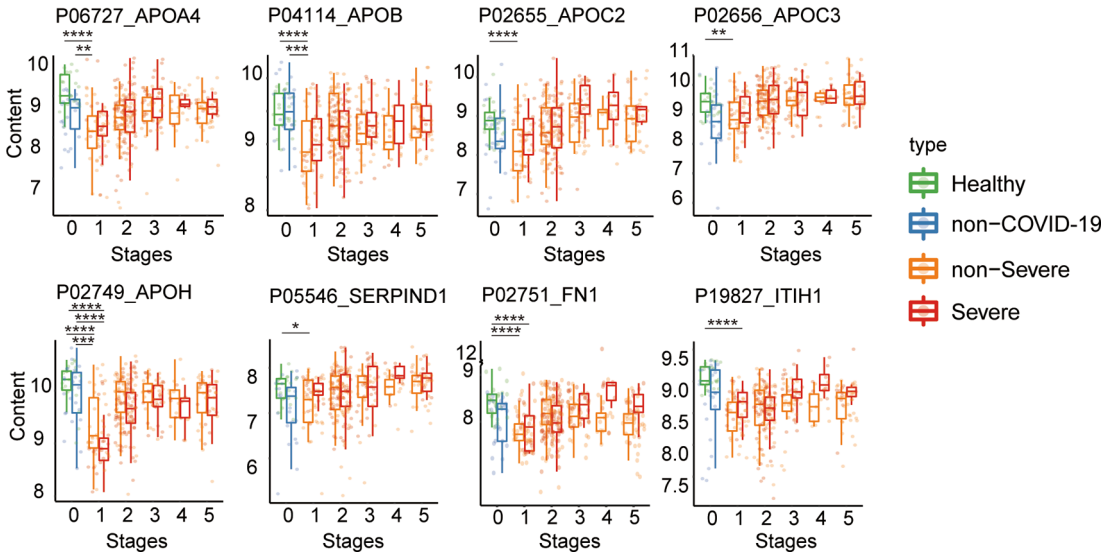


Figure 5. Eight significantly dysregulated proteins in COVID-19 compared to non-COVID-19 patients or healthy controls. Expression level changes of APOA4, APOB, APOC2, APOC3, APOH, SERPIND1, FN1, and ITIH1 in the five stages of the COVID-19 course as well as in non-COVID-19 patients and healthy controls. Asterisks indicate the statistical significance based on the unpaired two-sided Welch's *t* test p-value: *, < 0.05; **, < 0.01; ***, < 0.001; ****, < 0.0001.

For the non-severe group, another random forest model was built based on the serum proteome data (n = 108). This model also included 12 dysregulated proteins (Table S3C), as previously described, and 11 clinical indexes[37] with high accuracy (mean decrease accuracy > 0, which represents the loss of accuracy) (Figure 4B). The model achieved an AUC of 0.89 with the independent test set (Figure 4C), leading to the correct identification of 25 samples (prediction accuracy = 0.83) from the non-severe group (Figure 4D). The sample (CVDSBB611) from a 54 years old non-severe male patient was placed by our model in the NCP stage, even though it belonged to the NCN stage according to RT-PCR test. The nucleic acid test, however, was positive in his stool sample collected 14 days after discharge and negative in the pharyngeal swab. The complexity of the COVID-19 viral infection may thus provide several confounding factors. Another two samples (CVDSBB276 and CVDSBB278), incorrectly classified into the NCP stage, were both from a 53-year old non-severe female patient who underwent hysterectomy and was the only patient affected by anemia. Her surgery history and anemia status may thus have acted as confounding factors for our machine learning prediction. Additionally, one month after discharge, she was found positive for antinuclear antibodies, which may be another confounding factor. The relatively lower accuracy in predicting the disease course of non-severe patients was not a surprise: non-severe cases are generally less infected by the virus and tend to recover faster.

The selected features of the severe and the non-severe groups had only three overlapped proteins (APOH, APOE, and FN1) and one overlapped clinical index (the days after the symptoms' onset) between the two groups and thus supported the need to build two separate models.

We also attempted to build models for the COVID-19 patients using the 11 clinical indexes without proteins expression data. The prediction accuracy and AUC of the model were 0.80 and 0.85 for the 36 severe cases, respectively, while the accuracy and AUC of 108 non-severe cases were 0.77 and 0.86, respectively (Figure S5). We then used this model to classify those COVID-19 patients from whom specimens had been characterized by proteomics. The 34 non-severe cases achieved an accuracy of 0.77 and an AUC of 0.80, while the 19 severe cases exhibited a prediction accuracy of 0.90 and an AUC of 0.89 (Figure S6). Our models, including proteomics data, thus provided a higher prediction accuracy and AUC, further consolidating the value of serum proteins expression in COVID-19 monitoring.

Eight Significantly Dysregulated Proteins in COVID-19 Sera

On the basis of the dysregulated serum proteins shortlisted by our machine learning models, we further characterized the COVID-19 specific proteins by comparing their expression among three groups: stage 1 of COVID-19 patients, non-COVID-19 patients, and healthy controls. Comparing the stage 1 of COVID-19 patients with the healthy controls, we identified eight significantly dysregulated proteins, seven for the non-severe, and three for the severe subgroups (p-value < 0.05 and | log 2(fold change)| > log 2(1.2); Figure 5, Table S3D). These are apolipoprotein A-IV, B, C−II, C−III, H (APOA4, APOB, APOC2, APOC3, APOH), heparin cofactor 2 (SERPIND1), FN1, and ITIH1. Among them, compared with non-COVID-19 patients, APOB, APOA4, and APOH were also significantly dysregulated in the severe group, while APOH was dysregulated in the non-severe group. APOB and SERPIND1 have been reported to be associated with COVID-19 severity.[38]

Compared to the non-COVID-19 and the healthy groups, these eight proteins were all down-regulated at stage 1 and then gradually returned to relatively normal levels at stage 5 (Figure 5). APOA4, APOB, APOC2, and APOC3 are all apolipoproteins and components of the triglyceride-rich lipoproteins (TRLs) involved in lipid metabolism. Down-regulation of APOA4 in COVID-19 patients has been documented previously.[39] Among the TRLs, low-density lipoprotein (LDL) significantly decreases in COVID-19 patients and has been thus considered as a predictor of poor prognosis.[33] As the major component of LDL, APOB was downregulated at stage 1, possibly modulated by proinflammatory cytokines. During the infection, induced proinflammatory cytokines not only regulated the lipids metabolism but also increased the vascular permeability and

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

thus led to the plasma lipid leak.[33] The downregulation of these serum apolipoproteins in COVID-19 patients suggested a hyperinflammatory response and a damaged vascular permeability. APOH, SERPIND1, and FN1 all contributed to blood coagulation. APOH was reported to inhibit the aggregation of platelets and the intrinsic blood coagulation cascade.[40,41] SERPIND1, a protease inhibitor of the anticoagulant system, inhibits thrombin,[42] while thrombin is a coagulation factor of the extrinsic pathway of blood coagulation and could also induce the aggregation of platelets.[43] Plasma FN1 is an extracellular glycoprotein that participates in the formation of platelet clumps and stabilizes the aggregation of platelets.[44] The dysregulation of APOH, SERPIND1, and FN1 and their dynamic expression may thus indicate a dysregulation of the coagulation system during COVID-19. In addition, FN1, a major component of the extracellular matrices, could interact with viral proteins and mediate their entry.[45,46]

Limitations of This Study

This study analyzed the proteome of 320 serum samples from 54 COVID-19 patients and 45 control individuals. On the basis of the current situation in China, it is difficult to obtain an independent validation cohort of COVID-19 patients. More specimens from independent hospitals should be included in future studies to further validate our results. The number of samples in the NCP stage is about twice that in the NCN stage, which is not ideal for ordinary machine learning. For this method to be clinically applicable, the protein measurements should be implemented by targeted assays using SRM/MRM. Finally, mass spectrometry is currently less accessible than PCR. To fully implement the strategy described, systematic efforts to optimize and validate mass spectrometers for clinical environments will thus be required.

CONCLUSION

In this study, we present a new strategy to monitor the progression of COVID-19 that may complement RT-PCR-based nucleic acid tests. We show that serum protein biomarkers coupled with machine learning could be used to monitor the progression of COVID-19. As far as we know, this study is the first to apply serum proteomics to complement nucleic acid monitoring and predict the disease course of COVID-19. As we lack of an effective way to reduce the false negatives of routine RT-PCR tests, MS-based proteomics may play a complementary role in such clinical applications. Compared with ELISA, which detects proteins (including cytokines) using antibodies, MS-based methods require relatively small amounts of serum samples (1 μL), offer quantitative data, and can use high-temperature inactivated serum samples. Our analysis also identified 48 significantly dysregulated proteins during five stages of the COVID-19 progression.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.1c00525.

Details of SWATH-MS method (XLSX)
Clinical characteristics of 676 samples derived from 144 COVID-19 patients, 24 non-COVID-19 patients, and 21 healthy individuals (XLSX)
Proteomics data of 320 serum samples (XLSX)
Quality control of proteomics data acquired with SWATH-MS; biological replicates and technical replicates of proteomics data acquired with SWATH-MS; enriched pathways based on 48 significantly dysregulated proteins; Pearson correlation between protein expression and clinical indexes in severe and non-severe cases; machine learning using data from 144 patients; machine learning using data from 53 patients with proteomic data (PDF)

AUTHOR INFORMATION

Corresponding Authors

Hongguo Zhu − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*; Email: zhuhg@enzemed.com

Jiaqin Xu − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*; Email: xujq@enzemed.com

Yi Zhu − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*; Email: zhuyi@westlake.edu.cn

Tiannan Guo − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*; orcid.org/0000-0003-3869-7651; Email: guotiannan@westlake.edu.cn

Authors

Ying Zhang − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Xue Cai − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Weigang Ge − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China; Westlake Omics (Hangzhou) Biotechnology Co., Ltd., Hangzhou, Zhejiang 310000, China*

Donglian Wang − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Guangjun Zhu − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Liujia Qian − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious*

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China

Nan Xiang − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China; Westlake Omics (Hangzhou) Biotechnology Co., Ltd., Hangzhou, Zhejiang 310000, China*

Liang Yue − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Shuang Liang − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Fangfei Zhang − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Jing Wang − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Kai Zhou − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Yufen Zheng − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Minjie Lin − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Tong Sun − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Ruyue Lu − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Chao Zhang − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Luang Xu − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Yaoting Sun − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Xiaoxu Zhou − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Jing Yu − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Mengge Lyu − *Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310000, China; Center for Infectious Disease Research, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310000, China; Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310000, China*

Bo Shen − *Taizhou Hospital of Zhejiang Province Affiliated to Wenzhou Medical University, Linhai, Zhejiang 317000, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.1c00525

Author Contributions

[#]Ying Zhang, Xue Cai, Weigang Ge, Donglian Wang, Guangjun Zhu, and Liujia Qian are cofirst authors. T.G., Y. Zhang., X.C., and B.S. designed the project; Y. Zhang., D.W., and G.Z. performed the collection of samples and clinical data; X.C. and N.X. performed the sample preparation; X.C. performed SWATH-MS analysis; Y. Zhang., M.L., and T.S. took part in the acquisition of clinical data; R.L. organized data; X.C., W.G., F.Z., and X.Z. analyzed data; Y. Zhang., X.C., L.Q., B.S., T.G., and Y. Zhu wrote the manuscript with input from L.S., L.Y., L.X., J.Y., M.L., and Y.S.; C.Z. carried out the clinical supervision; T.G., Y. Zhu, H.Z., J.X., and B.S. supervised the project. All authors read and approved the final manuscript.

Notes

The authors declare the following competing financial interest(s): T.G. and Y.Z. are shareholders of Westlake Omics Inc. W.G. and N.X. are employees of Westlake Omics Inc. The remaining authors declare no competing interests.

The proteomics data and spectral library files in this study have been deposited to ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the iProX partner repository with the data set identifier PXD022751. All the data will be publicly released upon publication.

■ ACKNOWLEDGMENTS

This work was supported by grants from the National Key R&D Program of China (No. 2020YFE0202200), Zhejiang Province Analysis Test Project (2018C37032), the National Natural Science Foundation of China for Young Scholars (21904107), Zhejiang Provincial Natural Science Foundation for Distinguished Young Scholars (LR19C050001), Zhejiang Medical and Health Science and Technology Plan (2021KY394), Hangzhou Agriculture and Society Advancement Program (20190101A04), and Westlake Education Foundation, Tencent Foundation. We thank the Guomics team for helpful comments on this study, and the Westlake University Supercomputer

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

Center and biomedical research core facilities for assistance in data generation and storage.

REFERENCES

(1) 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.; China Novel Coronavirus, I.; Research, T. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382* (8), 727−733.

(2) Chan, J. F.; Yuan, S.; Kok, K. H.; To, K. K.; Chu, H.; Yang, J.; Xing, F.; Liu, J.; Yip, C. C.; Poon, R. W.; Tsoi, H. W.; Lo, S. K.; Chan, K. H.; Poon, V. K.; Chan, W. M.; Ip, J. D.; Cai, J. P.; Cheng, V. C.; Chen, H.; Hui, C. K.; Yuen, K. Y. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **2020**, *395* (10223), 514−523.

(3) Petersen, E.; Koopmans, M.; Go, U.; Hamer, D. H.; Petrosillo, N.; Castelli, F.; Storgaard, M.; Al Khalili, S.; Simonsen, L. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* **2020**, *20* (9), e238−e244.

(4) NHCPRC (National Health Commission of the PRC) (**2020**). Diagnosis and Treatment Protocol for COVID-19 (Trial Version 5). http://www.nhc.gov.cn/jkj/s3577/202002/a5d6f7b8c48c451c87dba14889b30147.shtml.

(5) Thevarajan, I.; Nguyen, T. H. O.; Koutsakos, M.; Druce, J.; Caly, L.; van de Sandt, C. E.; Jia, X.; Nicholson, S.; Catton, M.; Cowie, B.; Tong, S. Y. C.; Lewin, S. R.; Kedzierska, K. Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. *Nat. Med.* **2020**, *26* (4), 453−455.

(6) Yongchen, Z.; Shen, H.; Wang, X.; Shi, X.; Li, Y.; Yan, J.; Chen, Y.; Gu, B. Different longitudinal patterns of nucleic acid and serology testing results based on disease severity of COVID-19 patients. *Emerging Microbes Infect.* **2020**, *9* (1), 833−836.

(7) Wang, W.; Xu, Y.; Gao, R.; Lu, R.; Han, K.; Wu, G.; Tan, W. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* **2020**, *323* (18), 1843−1844.

(8) Wax, R. S.; Christian, M. D. Practical recommendations for critical care and anesthesiology teams caring for novel coronavirus (2019-nCoV) patients. *Can. J. Anaesth.* **2020**, *67* (5), 568−576.

(9) Tahamtan, A.; Ardebili, A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev. Mol. Diagn.* **2020**, *20* (5), 453−454.

(10) Pan, Y.; Long, L.; Zhang, D.; Yuan, T.; Cui, S.; Yang, P.; Wang, Q.; Ren, S. Potential False-Negative Nucleic Acid Testing Results for Severe Acute Respiratory Syndrome Coronavirus 2 from Thermal Inactivation of Samples with Low Viral Loads. *Clin. Chem.* **2020**, *66* (6), 794−801.

(11) Zhou, K.; Sun, Y.; Li, L.; Zang, Z.; Wang, J.; Li, J.; Liang, J.; Zhang, F.; Zhang, Q.; Ge, W.; Chen, H.; Sun, X.; Yue, L.; Wu, X.; Shen, B.; Xu, J.; Zhu, H.; Chen, S.; Yang, H.; Huang, S.; Peng, M.; Lv, D.; Zhang, C.; Zhao, H.; Hong, L.; Zhou, Z.; Chen, H.; Dong, X.; Tu, C.; Li, M.; Zhu, Y.; Chen, B.; Li, S. Z.; Guo, T. Eleven Routine Clinical Features Predict COVID-19 Severity Uncovered by Machine Learning of Longitudinal Measurements. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3640.

(12) Shen, B.; Yi, X.; Sun, Y.; Bi, X.; Du, J.; Zhang, C.; Quan, S.; Zhang, F.; Sun, R.; Qian, L.; Ge, W.; Liu, W.; Liang, S.; Chen, H.; Zhang, Y.; Li, J.; Xu, J.; He, Z.; Chen, B.; Wang, J.; Yan, H.; Zheng, Y.; Wang, D.; Zhu, J.; Kong, Z.; Kang, Z.; Liang, X.; Ding, X.; Ruan, G.; Xiang, N.; Cai, X.; Gao, H.; Li, L.; Li, S.; Xiao, Q.; Lu, T.; Zhu, Y.; Liu, H.; Chen, H.; Guo, T. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **2020**, *182* (1), 59−72.

(13) Fu, Y.; Li, Y.; Guo, E.; He, L.; Liu, J.; Yang, B.; Li, F.; Wang, Z.; Li, Y.; Xiao, R.; Liu, C.; Huang, Y.; Wu, X.; Lu, F.; You, L.; Qin, T.; Wang, C.; Li, K.; Wu, P.; Ma, D.; Sun, C.; Chen, G. Dynamics and Correlation Among Viral Positivity, Seroconversion, and Disease Severity in COVID-19: A Retrospective Study. *Ann. Intern. Med.* **2021**, *174* (4), 453−461.

(14) Jacobi, A.; Chung, M.; Bernheim, A.; Eber, C. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clin Imaging* **2020**, *64*, 35−42.

(15) Fu, Z.; Tang, N.; Chen, Y.; Ma, L.; Wei, Y.; Lu, Y.; Ye, K.; Liu, H.; Tang, F.; Huang, G.; Yang, Y.; Xu, F. CT features of COVID-19 patients with two consecutive negative RT-PCR tests after treatment. *Sci. Rep.* **2020**, *10* (1), 11548.

(16) Sinitcyn, P.; Tiwary, S.; Rudolph, J.; Gutenbrunner, P.; Wichmann, C.; Yilmaz, S.; Hamzeiy, H.; Salinas, F.; Cox, J. MaxQuant goes Linux. *Nat. Methods* **2018**, *15* (6), 401.

(17) Wichmann, C.; Meier, F.; Virreira Winter, S.; Brunner, A. D.; Cox, J.; Mann, M. MaxQuant. Live Enables Global Targeting of More Than 25,000 Peptides. *Mol. Cell Proteomics* **2019**, *18* (5), 982−994.

(18) 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* (5), 513−520.

(19) Chi, H.; Liu, C.; Yang, H.; Zeng, W.-F.; Wu, L.; Zhou, W.-J.; Wang, R.-M.; Niu, X.-N.; Ding, Y.-H.; Zhang, Y.; Wang, Z.-W.; Chen, Z.-L.; Sun, R.-X.; Liu, T.; Tan, G.-M.; Dong, M.-Q.; Xu, P.; Zhang, P.-H.; He, S.-M. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* **2018**, *36* (11), 1059−1061.

(20) Zeng, W. F.; Zhou, X. X.; Zhou, W. J.; Chi, H.; Zhan, J.; He, S. M. MS/MS Spectrum Prediction for Modified Peptides Using pDeep2 Trained by Transfer Learning. *Anal. Chem.* **2019**, *91* (15), 9724−9731.

(21) Zhu, T.; Zhu, Y.; Xuan, Y.; Gao, H.; Cai, X.; Piersma, S. R.; Pham, T. V.; Schelfhorst, T.; Haas, R.; Bijnsdorp, I. V.; Sun, R.; Yue, L.; Ruan, G.; Zhang, Q.; Hu, M.; Zhou, Y.; Van Houdt, W. J.; Le Large, T. Y. S.; Cloos, J.; Wojtuszkiewicz, A.; Koppers-Lalic, D.; Bottger, F.; Scheepbouwer, C.; Brakenhoff, R. H.; van Leenders, G.; Ijzermans, J. N. M.; Martens, J. W. M.; Steenbergen, R. D. M.; Grieken, N. C.; Selvarajan, S.; Mantoo, S.; Lee, S. S.; Yeow, S. J. Y.; Alkaff, S. M. F.; Xiang, N.; Sun, Y.; Yi, X.; Dai, S.; Liu, W.; Lu, T.; Wu, Z.; Liang, X.; Wang, M.; Shao, Y.; Zheng, X.; Xu, K.; Yang, Q.; Meng, Y.; Lu, C.; Zhu, J.; Zheng, J.; Wang, B.; Lou, S.; Dai, Y.; Xu, C.; Yu, C.; Ying, H.; Lim, T. K.; Wu, J.; Gao, X.; Luan, Z.; Teng, X.; Wu, P.; Huang, S.; Tao, Z.; Iyer, N. G.; Zhou, S.; Shao, W.; Lam, H.; Ma, D.; Ji, J.; Kon, O. L.; Zheng, S.; Aebersold, R.; Jimenez, C. R.; Guo, T. DPHL: A DIA Pan-human Protein Mass Spectrometry Library for Robust Biomarker Discovery. *Genomics, Proteomics Bioinf.* **2020**, *18* (2), 104−119.

(22) Ge, W.; Liang, X.; Zhang, F.; Xu, L.; Xiang, N.; Sun, R.; Liu, W.; Xue, Z.; Yi, X.; Wang, B.; Zhu, J.; Lu, C.; Zhan, X.; Chen, L.; Wu, Y.; Zheng, Z.; Gong, W.; Wu, Q.; Yu, J.; Ye, Z.; Teng, X.; Huang, S.; Zheng, S.; Liu, T.; Yuan, C.; Guo, T. Computational Optimization of Spectral Library Size Improves DIA-MS Proteome Coverage and Applications to 15 Tumors. *J. Proteome Res.* **2021**, DOI: 10.1021/acs.jproteome.1c00640.

(23) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24* (21), 2534−2536.

(24) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* **2014**, *32*, 219.

(25) Parker, S. J.; Rost, H.; Rosenberger, G.; Collins, B. C.; Malmström, L.; Amodei, D.; Venkatraman, V.; Raedschelders, K.; Van Eyk, J. E.; Aebersold, R. Identification of a Set of Conserved Eukaryotic Internal Retention Time Standards for Data-independent Acquisition Mass Spectrometry. *Mol. Cell Proteomics* **2015**, *14* (10), 2800−2813.

(26) 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* (7), 966−968.

(27) Sun, R.; Hunter, C.; Chen, C.; Ge, W.; Morrice, N.; Liang, S.; Zhu, T.; Yuan, C.; Ruan, G.; Zhang, Q.; Cai, X.; Yu, X.; Chen, L.; Dai,

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX

S.; Luan, Z.; Aebersold, R.; Zhu, Y.; Guo, T. Accelerated Protein Biomarker Discovery from FFPE Tissue Samples Using Single-Shot, Short Gradient Microflow SWATH MS. *J. Proteome Res.* **2020**, *19* (7), 2732−2741.
(28) Kumar, L.; Futschik, M. E. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2007**, *2* (1), 5−7.
(29) Messner, C. B.; Demichev, V.; Wendisch, D.; Michalick, L.; White, M.; Freiwald, A.; Textoris-Taube, K.; Vernardis, S. I.; Egger, A.-S.; Kreidl, M.; Ludwig, D.; Kilian, C.; Agostini, F.; Zelezniak, A.; Thibeault, C.; Pfeiffer, M.; Hippenstiel, S.; Hocke, A.; von Kalle, C.; Campbell, A.; Hayward, C.; Porteous, D. J.; Marioni, R. E.; Langenberg, C.; Lilley, K. S.; Kuebler, W. M.; Mülleder, M.; Drosten, C.; Suttorp, N.; Witzenrath, M.; Kurth, F.; Sander, L. E.; Ralser, M. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Systems* **2020**, *11* (1), 11−24.
(30) D'Alessandro, A.; Thomas, T.; Dzieciatkowska, M.; Hill, R. C.; Francis, R. O.; Hudson, K. E.; Zimring, J. C.; Hod, E. A.; Spitalnik, S. L.; Hansen, K. C. Serum Proteomics in COVID-19 Patients: Altered Coagulation and Complement Status as a Function of IL-6 Level. *J. Proteome Res.* **2020**, *19* (11), 4417−4427.
(31) Dong, Y.; Zhou, H.; Li, M.; Zhang, Z.; Guo, W.; Yu, T.; Gui, Y.; Wang, Q.; Zhao, L.; Luo, S.; Fan, H.; Hu, D. A novel simple scoring model for predicting severity of patients with SARS-CoV-2 infection. *Transboundary Emerging Dis.* **2020**, *67* (6), 2823−2829.
(32) Taha, M.; Sano, D.; Hanoudi, S.; Esber, Z.; Elahi, M.; Gabali, A.; Chopra, T.; Draghici, S.; Samavati, L. Platelets and renal failure in the SARS-CoV-2 syndrome. *Platelets* **2021**, *32*, 1−8.
(33) Fan, J.; Wang, H.; Ye, G.; Cao, X.; Xu, X.; Tan, W.; Zhang, Y. Letter to the Editor: Low-density lipoprotein is a potential predictor of poor prognosis in patients with coronavirus disease 2019. *Metab., Clin. Exp.* **2020**, *107*, 154243.
(34) Ren, H.; Yang, Y.; Wang, F.; Yan, Y.; Shi, X.; Dong, K.; Yu, X.; Zhang, S. Association of the insulin resistance marker TyG index with the severity and mortality of COVID-19. *Cardiovasc. Diabetol.* **2020**, *19* (1), 58.
(35) Shu, T.; Ning, W.; Wu, D.; Xu, J.; Han, Q.; Huang, M.; Zou, X.; Yang, Q.; Yuan, Y.; Bie, Y.; Pan, S.; Mu, J.; Han, Y.; Yang, X.; Zhou, H.; Li, R.; Ren, Y.; Chen, X.; Yao, S.; Qiu, Y.; Zhang, D. Y.; Xue, Y.; Shang, Y.; Zhou, X. Plasma Proteomics Identify Biomarkers and Pathogenesis of COVID-19. *Immunity* **2020**, *53* (5), 1108−1122.
(36) Zheng, X.; Chen, J.; Deng, L.; Fang, Z.; Chen, G.; Ye, D.; Xia, J.; Hong, Z. Risk factors for the COVID-19 severity and its correlation with viral shedding: A retrospective cohort study. *J. Med. Virol.* **2021**, *93* (2), 952−961.
(37) Zhou, K.; Sun, Y.; Li, L.; Zang, Z.; Wang, J.; Li, J.; Liang, J.; Zhang, F.; Zhang, Q.; Ge, W.; Chen, H.; Sun, X.; Yue, L.; Wu, X.; Shen, B.; Xu, J.; Zhu, H.; Chen, S.; Yang, H.; Huang, S.; Peng, M.; Lv, D.; Zhang, C.; Zhao, H.; Hong, L.; Zhou, Z.; Chen, H.; Dong, X.; Tu, C.; Li, M.; Zhu, Y.; Chen, B.; Li, S. Z.; Guo, T., Eleven Routine Clinical Features Predict COVID-19 Severity. *medRxiv*, July 29, **2020**. DOI: 10.1101/2020.07.28.20163022.
(38) Suvarna, K.; Biswas, D.; Pai, M. G. J.; Acharjee, A.; Bankar, R.; Palanivel, V.; Salkar, A.; Verma, A.; Mukherjee, A.; Choudhury, M.; Ghantasala, S.; Ghosh, S.; Singh, A.; Banerjee, A.; Badaya, A.; Bihani, S.; Loya, G.; Mantri, K.; Burli, A.; Roy, J.; Srivastava, A.; Agrawal, S.; Shrivastav, O.; Shastri, J.; Srivastava, S. Proteomics and Machine Learning Approaches Reveal a Set of Prognostic Markers for COVID-19 Severity With Drug Repurposing Potential. *Front. Physiol.* **2021**, *12*, 652799.
(39) Di, B.; Jia, H.; Luo, O. J.; Lin, F.; Li, K.; Zhang, Y.; Wang, H.; Liang, H.; Fan, J.; Yang, Z. Identification and validation of predictive factors for progression to severe COVID-19 pneumonia by proteomics. *Signal Transduct Target Ther* **2020**, *5* (1), 217.
(40) Nimpf, J.; Wurm, H.; Kostner, G. M. Beta 2-glycoprotein-I (apo-H) inhibits the release reaction of human platelets during ADP-induced aggregation. *Atherosclerosis* **1987**, *63* (2−3), 109−14.
(41) Schousboe, I. beta 2-Glycoprotein I: a plasma inhibitor of the contact activation of the intrinsic blood coagulation pathway. *Blood* **1985**, *66* (5), 1086−91.
(42) Iumashev, G. S.; Kriukov, B. N.; Rappoport, N. [Results of treatment of patients with trochanteric fractures of the femoral bone]. *Sov Med.* **1977**, No. 3, 37−41.
(43) van 't Veer, C.; Mann, K. G. Regulation of tissue factor initiated thrombin generation by the stoichiometric inhibitors tissue factor pathway inhibitor, antithrombin-III, and heparin cofactor-II. *J. Biol. Chem.* **1997**, *272* (7), 4367−77.
(44) Cho, J.; Mosher, D. F. Role of fibronectin assembly in platelet thrombus formation. *J. Thromb. Haemostasis* **2006**, *4* (7), 1461−9.
(45) He, Q. Q.; Ren, S.; Xia, Z. C.; Cheng, Z. K.; Peng, N. F.; Zhu, Y. Fibronectin Facilitates Enterovirus 71 Infection by Mediating Viral Entry. *J. Virol.* **2018**, *92* (9), e02251.
(46) Proctor, R. A. Fibronectin: a brief overview of its structure, function, and physiology. *Clin. Infect. Dis.* **1987**, *9* (Suppl 4), S317−21.

https://doi.org/10.1021/acs.jproteome.1c00525
J. Proteome Res. XXXX, XXX, XXX−XXX