Received: 8 November 2020 | Revised: 25 March 2021 | Accepted: 30 March 2021

DOI: 10.1002/pmic.202000278

Proteomics
Proteomics and Systems Biology

RESEARCH ARTICLE

Longitudinal proteomic profiling provides insights into host response and proteome dynamics in COVID-19 progression

Jee-Soo Lee[1] | Dohyun Han[2] | So Yeon Kim[3] | Ki Ho Hong[4] | Myoung-jin Jang[5] | Man Jin Kim[1] | Young-gon Kim[1] | Jae Hyeon Park[1] | Sung Im Cho[1] | Wan Beom Park[6] | Kyung Bok Lee[7] | Ho Seob Shin[1] | Hyeon Sae Oh[1] | Taek Soo Kim[1] | Sung Sup Park[1] | Moon-Woo Seong[1]

[1] Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, South Korea

[2] Proteomics Core Facility, Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea

[3] Department of Laboratory Medicine, National Medical Center, Seoul, South Korea

[4] Department of Laboratory Medicine, Seoul Medical Center, Seoul, South Korea

[5] Medical Research Collaborating Center, Seoul National University Hospital

[6] Department of Internal Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, South Korea

[7] Department of Statistics, Seoul National University, Seoul, South Korea

Correspondence
Moon-Woo Seong, Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea.
Email: mwseong@snu.ac.kr

Jee-Soo Lee, Dohyun Han, So Yeon Kim, and Ki Ho Hong contributed equally to this work.

Abstract

In managing patients with coronavirus disease 2019 (COVID-19), early identification of those at high risk and real-time monitoring of disease progression to severe COVID-19 is a major challenge. We aimed to identify potential early prognostic protein markers and to expand understanding of proteome dynamics during clinical progression of the disease. We performed in-depth proteome profiling on 137 sera, longitudinally collected from 25 patients with COVID-19 (non-severe patients, $n = 13$; patients who progressed to severe COVID-19, $n = 12$). We identified 11 potential biomarkers, including the novel markers IGLV3-19 and BNC2, as early potential prognostic indicators of severe COVID-19. These potential biomarkers are mainly involved in biological processes associated with humoral immune response, interferon signalling, acute phase response, lipid metabolism, and platelet degranulation. We further revealed that the longitudinal changes of 40 proteins persistently increased or decreased as the disease progressed to severe COVID-19. These 40 potential biomarkers could effectively reflect the clinical progression of the disease. Our findings provide some new insights into host response to SARS-CoV-2 infection, which are valuable for understanding of COVID-19 disease progression. This study also identified potential biomarkers that could be further validated, which may support better predicting and monitoring progression to severe COVID-19.

KEYWORDS
Keywords, COVID-19, proteomics, severity, prognosis, serum

1 | INTRODUCTION

A wide range of disease severity, ranging from asymptomatic to life-threatening illness, has been observed in patients with coronavirus disease 2019 (COVID-19) [1]. While the current estimate is that 81% of COVID-19 patients have a mild or moderate disease course, some patients who initially present with mild symptoms subsequently experience acute clinical deterioration within 1–2 weeks after symptom

Abbreviations: COVID-19, coronavirus disease 2019; LDH, lactate dehydrogenase; CRP, C-reactive protein; IFN, interferon; TNF, tumour necrosis factor; MS, mass spectrometry; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; DEP, differentially expressed protein; TFA, trifluoroacetic acid; SDB-RPS, styrene divinylbenzene reversed-phase sulfonate; LC-MS/MS, Liquid chromatography-tandem mass spectrometry; DIA, data-independent acquisition; DDA, data-dependent acquisition; FDR, false discovery rate; GO, gene ontology; CI, confidence interval; BCV, best cut-off value; IQR, interquartile range; PCC, Pearson's correlation coefficients; ACE2, angiotensin I-converting enzyme 2; LPS, lipopolysaccharides; TLR4, toll-like receptor 4; MD-2, myeloid differentiation factor 2

onset [1, 2]. Therefore, precise prognostication and early identification of disease progression are needed to effectively manage patients who are at high risk of clinical progression to severe COVID-19.

Several previous studies have reported clinical laboratory findings including lactate dehydrogenase (LDH), C-reactive protein (CRP), albumin, D-dimer, lymphocytes, neutrophils, and platelets at admission as parameters for progressive risk estimation [3–6]. Plasma levels of proteins associated with cytokine storms and immune response including interleukin (IL)-2, IL-6, IL-7, G-CSF, interferon (IFN)-γ inducible protein 10, monocyte chemoattractant protein-1, macrophage inflammatory protein-1α, and tumour necrosis factor (TNF)-α have been reported to be elevated in severe COVID-19 patients [2, 7, 8]. In addition, mass spectrometry (MS)-based quantitative proteome profiling has been performed in several recent studies, which reveal characteristic molecular changes and biological processes in patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and at high risk of clinical progression [9, 10]. These studies have reported that severe COVID-19 patients have altered abundance of proteins involved in immune/inflammatory regulation (e.g., IL-6 and type 1 IFN), coagulation (e.g., fibrinogen and SERPINA10), lipid metabolism, and acute phase response (e.g., CRP) [11, 12]. Further proteome profiling may enable us to expand the understanding of prognostic molecular signatures altered at initial assessment or during the disease course in patients with a high risk of disease progression.

In this study, we addressed two questions about prognostication in COVID-19: (1) Which protein markers are differentially expressed preceding clinical deterioration and are able to significantly predict the high risk in patients who progressed to severe COVID-19 compared to patients who had a non-severe disease course. (2) Which differentially expressed protein (DEP) markers exhibit longitudinal changes that effectively reflect clinical progression of the disease. We performed longitudinal proteome profiling in serum samples of COVID-19 patients at admission, and serially collected samples during their hospital stay. We then applied pathway analysis to gain insight into the relevant biological processes of altered protein expression. The overall strategy in this study is shown in Figure 1.

Significance Statement

The coronavirus disease 2019 (COVID-19) pandemic highlights the need for biomarkers to identify the patients at high risk of severe COVID-19 early and to monitor the disease progression. In this study, we performed longitudinal proteomic profiling on sera serially collected from patients with COVID-19. Findings from this study identified 11 potential early prognostic indicators, including the novel protein markers IGLV3-19 and BNC2. IGLV3-19 is a variable domain of immunoglobulin light chains, which may provide better understanding of the host immune response in severe COVID-19. BNC2 might play a role in the upregulation of interferon-stimulated genes in the interferon signalling pathway. We further identified that the longitudinal changes of 40 proteins persistently increased or decreased as the disease progressed to severe COVID-19, which effectively reflect the clinical progression of the disease. This study identified potential biomarkers that could be further validated, which may support better predicting and monitoring progression to severe COVID-19.

2 | MATERIALS AND METHODS

2.1 | Patients and samples

A total of 137 serum samples (sampled at admission, $n = 25$; longitudinally sampled during hospitalization, $n = 112$) were collected from 25 patients with laboratory-confirmed SARS-CoV-2 infection (Figure S1). Each sampling time point is described in Table S1.

Patients were from two tertiary hospitals in South Korea (National Medical Center and Seoul Medical Center). The patients' clinical status was assessed daily during hospitalization, and they were classified into two groups, namely non-severe patients and patients who progressed to severe COVID-19, according to the 8-point ordinal scale (defined in Table S2) as described by Messner et al. [12, 13]. This study was approved by the institutional review board (IRB No. H-2004-153-1118).

2.2 | Sample preparation

The protein digestion process was optimized to 2 μL of serum as previously described, with some modifications [14]. Briefly, 23 μL of digestion buffer [8 M urea, 10 mM Tris (2-carbocyethyl) phosphine, and 50 mM chloroacetamide in 50 mM ammonium bicarbonate] was added to 2 μL of serum. The mixture was incubated for 25 min at 60°C for simultaneous denaturation, reduction, and alkylation of proteins. After cooling to room temperature, protein digestion was performed at 37°C overnight using a trypsin/LysC mixture at a 100:1 protein-to-protease ratio. The second digestion was performed at 37°C for 2 h using trypsin (enzyme-to-substrate ratio [w/w] of 1:1000). All resulting peptides were acidified with 10% trifluoroacetic acid (TFA). The acidified peptides were loaded onto a custom styrenedivinylbenzene reversed-phase sulfonate (SDB-RPS)-StageTips, as described previously [14]. Following washing with 0.2% TFA, the peptides were eluted with 80% acetonitrile containing 1% ammonia. The eluate was dehydrated in a vacuum-centrifuge and stored at -80°C.

2.3 | Liquid chromatography-tandem mass spectrometry (LC-MS/MS) and data analysis

All LC-MS/MS analyses were performed using Quadrupole Orbitrap mass spectrometers, Q-Exactive Plus (Thermo Fisher Scientific,

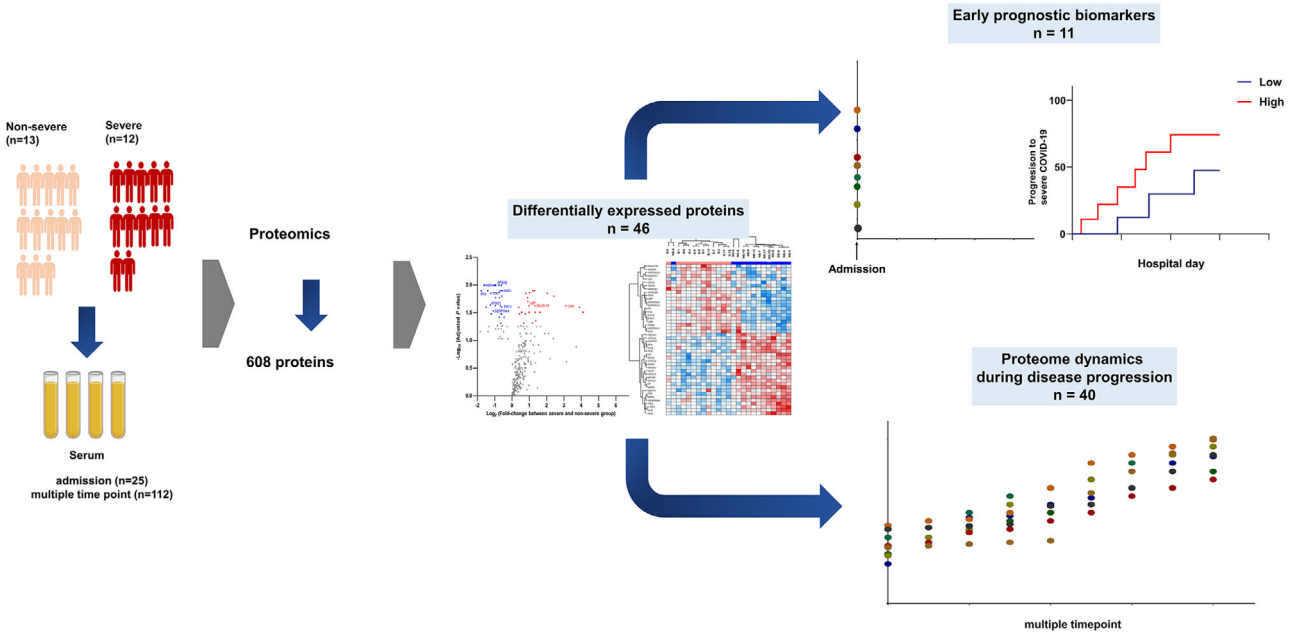

FIGURE 1 Study workflow. In-depth proteome profiling was performed on 137 serum samples collected from 25 patients with COVID-19 (non-severe, $n = 13$; severe, $n = 12$). First, differentially expressed proteins (DEPs) between the two groups were identified. Then, the abundance of each candidate protein at admission and the risk probability of progression to severe COVID-19 were evaluated. Also, dynamics of DEPs in patients with disease progression to severe COVID-19 was assessed.

Waltham, MA, USA) coupled to an Ultimate 3000 RSLC system (Dionex, Sunnyvale, CA, USA) via a nano electrospray source, as described with some modifications [15]. Peptide samples were separated on a 2-column setup, with a trap column (75 μm I.D. × 2 cm, C18 3 μm, 100 Å) and an analytical column (50 μm I.D. × 15 cm, C18 1.9 μm, 100 Å). Prior to sample injection, the dried peptide samples were redissolved in solvent A (2% acetonitrile and 0.1% formic acid). After the samples were loaded onto the nano LC, a 90-min gradient from 8% to 26% solvent B (100% acetonitrile and 0.1% formic acid) was applied to all samples. The spray voltage was 2.0 kV in the positive ion mode, and the temperature of the heated capillary was set to 320°C. The data-independent acquisition (DIA) method consisted of a survey scan at 35,000 resolution from 400 to 1220 m/z (AGC target of 3×10^6 or 60 ms injection time). Then, 19 DIA windows were acquired at 35,000 resolutions with an automatic gain control target of 3×10^6 and auto for injection time [16]. The stepped collision energy was 10% at 27%.

2.4 | Spectral library generation

To generate the spectral libraries for DIA, 24 data-dependent acquisition (DDA) measurements of the immunodepleted plasma samples were performed. DDA spectra were searched using the Maxquant against Uniprot Human Database (December 2014, 88,657 entries) and the iRT standard peptide sequence. A spectral library for DIA-MS was generated using spectral library generation in Spectronaut PulsarX (Biognosys, Schlieren, Switzerland). In addition, we used Spectronaut Pulsar's protein identification algorithm for spectral library generation. For the generation of Spectronaout Pulsar spectral libraries, DIA raw data were directly loaded into Spectronaut Pulsar, and spectral library generation was performed using the default settings

2.5 | Data processing for DIA-MS

The DIA data of individual samples were analysed with Spectronaut 13 (Biognosys, Schlieren, Switzerland). First, the DIA raw files were converted into an htrm format using the GTRMS Converter provided by Spectronaut. The false discovery rate (FDR) was estimated with the mProphet [17] approach and set to 1% at the peptide precursor level and 1% at the protein level. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [18] partner repository with the dataset identifier PXD023686. The proteins were inferred by the software, and the quantification information was acquired at the protein level by using the q-value < 0.01 criteria, which was used for subsequent analyses. The relative abundance of all proteins in all longitudinal samples were presented in Table S3.

2.6 | Quantification and statistical analysis

2.6.1 | Differential expression analysis

Serum protein levels were estimated using the protein group quantity calculated using Spectronaut software. Proteins with > 50% missing ratios in a particular patient group were removed. Following $\log_2$ transformation, missing values were imputed based on a normal distribution

(width = 0.3, downshift = 1.8) to simulate signals of low-abundance proteins. Fold change was calculated using the mean of a particular patient group. DEPs between the two groups were identified using the Mann-Whitney U test. The level of statistical significance was set as adjusted p value < 0.05, and the FDR was corrected using Benjamini and Hochberg for multiple testing. Protein abundances were subjected to z-normalization followed by hierarchical clustering with Pearson's correlation distance.

2.6.2 | Pathway analysis

Canonical pathways, disease and functions, and protein networks were evaluated by Ingenuity Pathway Analysis (IPA, QIAGEN, Hilden, Germany) based on the DEPs in the serum. The analytical algorithms embedded in IPA uses lists of DEPs to predict biological processes and pathways. Additionally, a tree map was constructed, in which the major boxes represented categories of related biological functions or diseases. The statistical significance of both the gene ontology (GO) classification and enrichment analysis was determined by Fisher's exact test. All statistical tests were two-sided, and p value < 0.05 was considered statistically significant. Wiki pathway and GO biological processes were identified using the Enrichr online tool (http://amp.pharm.mssm.edu/Enrichr/). The top ten enriched terms (and their p values) were identified.

2.6.3 | Statistical analysis

We evaluated the association between the level of each candidate protein marker (i.e., DEPs) at admission and estimated the risk of disease progression using the Cox proportional hazard model. Hazard ratios with 95% confidence intervals (CIs) were calculated using Cox proportional hazard regression. The best cut-off value (BCV) of each candidate protein was determined using maximally selected log-rank statistics, as previously described [19]. The time to clinical progression to severe COVID-19 was plotted using Kaplan–Meier curves, and compared with a log-rank test; p value < 0.05 was considered statistically significant. We further explored the effect of longitudinal changes in each candidate protein level on the risk of clinical progression using a linear mixed-effects model; the level of statistical significance was set as adjusted p value < 0.05, FDR corrected [20]. Analysis was performed using commercially available software (SPSS 19.0, Stata SE12.0, and SAS version 9.4).

3 | RESULTS

3.1 | Clinical characteristics of patients

The demographics and clinical characteristics of the study patients are summarized in Table 1 and Figure S2. The median age was 70 years (interquartile range [IQR], 65 to 76 years), and 48.0% of the patients were male. The median time from symptom onset to admission was 5 days (IQR, 4–9 days). All patients were followed up for >10 days after admission. Among the total population, 13 (52.0%) patients who maintained an ordinal scale ≤ 4 (category 3, n = 4; category 4, n = 9) were classified into the non-severe group, and 12 (48.0%) patients with clinical deterioration to ordinal scale >4 (category 5, n = 9; category 6, n = 2; category 8, n = 1) were classified into the severe group. In the non-severe group, 9 (69.2%) patients required supplemental oxygen using either nasal prongs or mask; all the patients in the non-severe group improved and were discharged. In the severe group, the median time from admission to clinical deterioration requiring high-flow oxygen was 5 days (IQR, 3.5- 7.5 days). Invasive mechanical ventilation and renal replacement therapy were administered in 25.0% (n = 3) and 8.3% (n = 1) of patients in the severe group, respectively. One patient (8.3%) died, and 11 (91.7%) patients were discharged with clinical improvement.

3.2 | DEPs in severe COVID-19 patient serum

Overall, 608 proteins were identified and quantified in the cross-sectional samples. An average of 484 proteins were quantified in the individual samples (Figure S3A). To identify the differences within and between groups, protein profiles were generated by drawing multi-scatter plots and calculating Pearson's correlation coefficients (PCCs; Figure S3B). Non-severe and severe groups showed mean PCCs of 0.88 and 0.89, respectively. The average PCC between the non-severe and severe groups was 0.87. Based on serum protein expression levels of patients in the non-severe and severe groups, two groups were separated by principal component analysis (Figure S3C). After proteins were determined to be quantified by at least 50% in either the non-severe or severe group subjects, 513 proteins were subjected to statistical analysis.

We found that 46 proteins were differentially expressed in the serum, sampled immediately after patients had progressed to severe COVID-19, in severe patients compared with that in non-severe patients (Table 2, Figure 2A). Pathway analysis and GO enrichment analysis revealed that the DEPs were mainly involved in several biological processes associated with humoral immune response (IGLV3-19, IGLC2, and IGHA2), IFN signalling (BNC2), acute phase response (CRP, LBP, SERPINA3, SAA1, SAA2, ITIH4, RBP4, TTR, albumin, and transferrin), inflammatory response (angiotensinogen, HSP90AA1, transketolase), lipid metabolism (APOA1, APOA2, APOA4, APOC1, APOM, and paraoxonase-1), platelet degranulation (LEFTY2, AHSG, SEPP1, A2M, KNG1, PF4, HRG, and SERPINA4), coagulation cascade (FIX, FX, SERPINA1, and SERPING1), and cellular metabolic process (ALDOA, PIK3C2β, MAN1A, MAN1C1, Gc-globulin, ITIH2, IGFALS, and PI16) (Figure 2B, Figure S4).

TABLE 1 Demographics and clinical characteristics of COVID-19 patients

	Total ($N = 25$)	Non-severe ($n = 13$)	Severe ($n = 12$)
Age – year (median, IQR)	70 (65–76)	71 (64–77)	69.5 (67.3–72)
Male sex – No. (%)	12 (48)	4 (30.8)	8 (66.7)
Time from symptom onset to admission	5 (4–9)	5 (4–14)	4 (4–8)
Eight-point score on ordinal scale – No (%)			
3: hospitalized, not requiring supplemental oxygen	4 (16.0)	4 (30.8)	0 (0)
4: hospitalized, requiring low-flow oxygen by nasal prongs or facial mask	9 (36.0)	9 (69.2)	0 (0)
5: hospitalized, use of high-flow oxygen devices or non-invasive ventilation	9 (36.0)	0 (0)	9 (75.0)
6: hospitalized, intubation and invasive mechanical ventilation	2 (8.0)	0 (0)	2 (16.7)
7: hospitalized, invasive mechanical ventilation, and ECMO, RRT or both	0 (0)	0 (0)	0 (0)
8: death	1 (4.0)	0 (0)	1 (8.3)
Time from admission to severe disease			5 (3.5–7.5)
Symptoms			
Fever	10 (40.0)	5 (38.5)	5 (41.7)
Chill	4 (16.0)	1 (7.7)	3 (25.0)
Cough	8 (32.0)	2 (15.4)	6 (50.0)
Sputum	6 (24.0)	2 (15.4)	4 (33.3)
Dyspnoea	3 (12.0)	1 (7.7)	2 (16.7)
Myalgia	6 (24.0)	2 (15.4)	4 (33.3)
Fatigue	1 (4.0)	1 (7.7)	0 (0)
Headache	2 (8.0)	1 (7.7)	1 (8.3)
Rhinorrhoea	1 (4.0)	1 (7.7)	0 (0)
Sore throat	4 (16.0)	2 (15.4)	2 (16.7)
Epigastric pain	1 (4.0)	0 (0)	1 (8.3)
Diarrheal	2 (8.0)	1 (7.7)	1 (8.3)
Indigestion	1 (4.0)	0 (0)	1 (8.3)
Olfactory and gustatory sensory dysfunction	2 (8.0)	1 (7.7)	1 (8.3)
Comorbidity			
HTN	10 (40.0)	6 (46.2)	4 (33.3)
DM	6 (24.0)	4 (30.8)	2 (16.7)
Dyslipidaemia	4 (16.0)	2 (15.4)	2 (16.7)
Dementia	4 (16.0)	2 (15.4)	2 (16.7)
Parkinson's disease	1 (4.0)	0 (0)	1 (8.3)
Treatment			
Lopinavir-Ritonavir	8 (32.0)	4 (30.8)	4 (33.3)
Antibiotic agent	2 (8.0)	0 (0)	2 (16.7)
Glucocorticoid therapy	2 (8.0)	0 (0)	2 (16.7)
Oxygen by nasal prongs or mask	21 (84.0)	9 (69.2)	12 (100)
High flow oxygen or non-invasive mechanical ventilation	12 (48.0)	0 (0)	12 (100)
Invasive mechanical ventilation	3 (12.0)	0 (0)	3 (25.0)
Renal replacement therapy	1 (4.0)	0 (0)	1 (8.3)
ECMO	0 (0)	0 (0)	0 (0)
Outcomes at discharge			
Improved	24 (96.0)	13 (100)	11 (91.7)
Died	1 (4.0)	0 (0)	1 (8.3)

Abbreviations: IQR, interquartile range; ECMO, extracorporeal membrane oxygenation; RRT, renal replacement therapy; HTN, hypertension; DM, diabetes mellitus.

TABLE 2 Differentially expressed proteins between non-severe and severe COVID-19 patient groups

Protein	Gene	FC	*p* value	adjusted *p* value
A0A075B6J8	IGLV3-19	1.3167	0.00146	0.02345
A0A075B6K9	IGLC2	1.0179	0.00045	0.01361
A0A087 × 1L7	LEFTY2	−0.6119	0.00013	0.01004
B1APH0	BNC2	−0.5972	0.00176	0.02512
D6RAR4	HGFAC	−0.6465	0.00509	0.04836
D6REX5	SEPP1	−0.4608	0.00361	0.03784
D6RF35	GC	−0.6678	0.00146	0.02345
K7ERI9	APOC1	−1.2537	0.00121	0.023
O95445	APOM	−0.6069	0.00303	0.0331
P00740	F9	0.6716	0.00509	0.04836
P00742	F10	0.3891	0.00176	0.02512
P01009	SERPINA1	0.9094	0.001	0.01971
P01011	SERPINA3	1.286	0.00024	0.01255
P01019	AGT	0.5587	0.00254	0.031
P01023	A2M	−0.7402	0.00361	0.03784
P01042	KNG1	−0.5741	0.00037	0.01255
P01877	IGHA2	0.9939	0.00254	0.031
P02647	APOA1	−0.7496	0.00016	0.01004
P02741	CRP	3.1016	0.00146	0.02345
P02753	RBP4	−1.615	0.00016	0.01004
P02765	AHSG	−1.0288	0.00008	0.01004
P02766	TTR	−1.232	0.00055	0.01417
P02768	ALB	−0.9485	0.00082	0.01687
P02776	PF4	−1.7738	0.00037	0.01255
P02787	TF	−0.9503	0.00013	0.01004
P04075	ALDOA	1.6074	0.00254	0.031
P04196	HRG	−0.9778	0.00146	0.02345
P05155	SERPING1	0.4239	0.00303	0.0331
P06727	APOA4	−1.2004	0.00303	0.0331
P07900	HSP90AA1	1.173	0.00509	0.04836
P0DJI8	SAA1	3.8954	0.00176	0.02512
P0DJI9	SAA2	4.1052	0.00254	0.031
P18428	LBP	0.9799	0.00146	0.02345
P19823	ITIH2	−0.5939	0.00067	0.01574
P27169	PON1	−0.6905	0.00037	0.01255
P29401	TKT	1.3681	0.0043	0.04407
P29622	SERPINA4	−1.0806	0.00212	0.02936
P33908	MAN1A1	1.3383	0.00254	0.031
P35858	IGFALS	−0.7291	0.00082	0.01687
Q14624	ITIH4	2.4059	0.00067	0.01574
Q5SWA0	PIK3C2B	2.0339	0.00055	0.01417
Q5T7N2	L1TD1	−0.6234	0.00303	0.0331
Q6UXB8	PI16	−1.4981	0.00176	0.02512
Q92954	PRG4	1.1984	0.00037	0.01255

(Continues)

TABLE 2 (Continued)

Protein	Gene	FC	*p* value	adjusted *p* value
Q9NR34	MAN1C1	7.7997	0.00508	0.04836
V9GYM3	APOA2	−0.9625	0.00004	0.01004

Abbreviation: FC, fold-change (Log2 fold change was calculated.)
Samples obtained immediately after patients progressed to severe COVID-19 (i.e., high-flow O2 start) for the severe COVID-19 group and samples at admission for non-severe COVID-19 group were assessed.

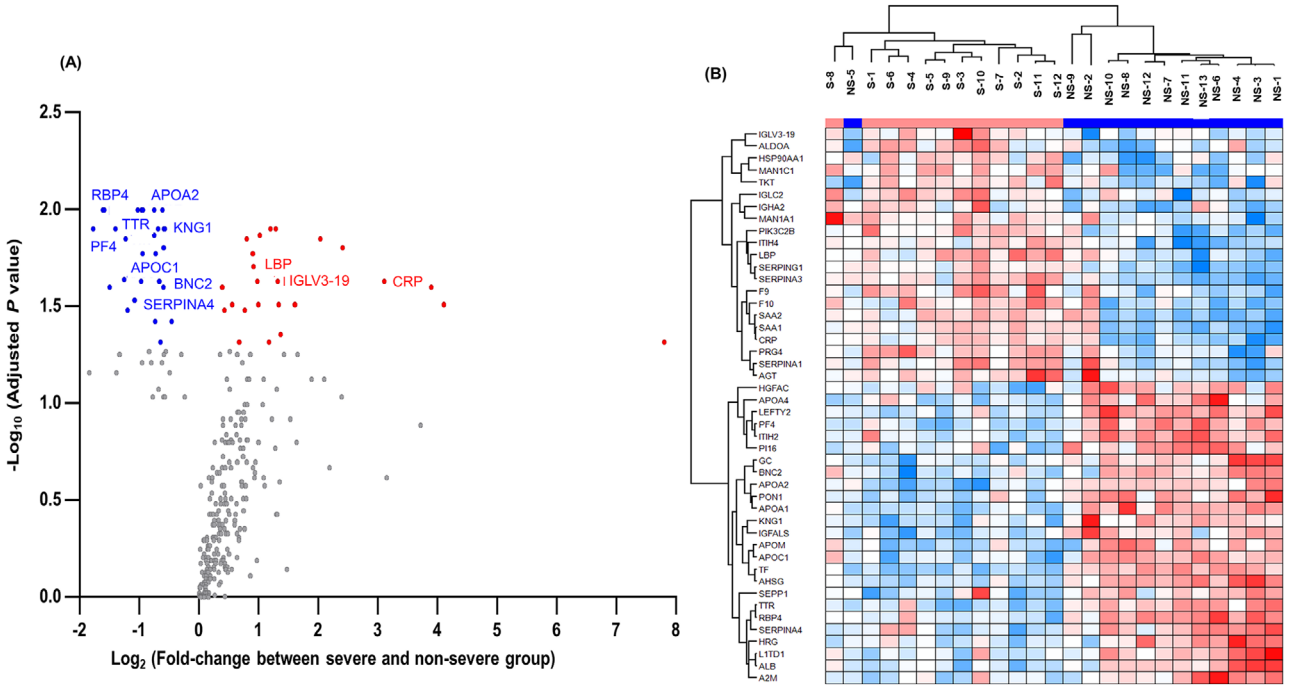

FIGURE 2 (A) Volcano plot analysis of differentially expressed proteins (DEPs) between two groups. The expression difference in protein between severe groups and non-severe groups is plotted on the x-axis, and false discovery rate – adjusted p value is plotted on the y-axis (-log10 scale). Upregulated and downregulated proteins in the severe group are represented red and blue, respectively. (B) Gene ontology enrichment analysis. Heatmap of 46 proteins that are differentially abundant in severe COVID-19. Protein levels are presented on a low to high scale (blue – white – red).

3.3 | Early prognostic biomarkers for progression to severe COVID-19

We revealed that several DEPs (IGLV3-19, BNC2, CRP, LBP, RBP4, TTR, APOC1, APOA2, KNG1, PF4, and SERPINA4) were candidate predictors of progression to severe COVID-19. Notably, IGLV3-19, a variable domain of immunoglobulin light chains, and BNC2, an activator of genes in the IFN signalling pathway [21], are novel candidate markers that have not yet been reported in SARS-CoV-2 infection.

Figure 3 shows the significant linear correlation between the abundance of each protein at admission and the risk probability of progression to severe COVID-19. The levels of IGLV3-19, CRP, LBP, and the levels of BNC2, RBP4, TTR, APOC1, APOA2, KNG1, PF4, and SERPINA4 at admission were positively and negatively correlated with the risk of clinical deterioration to severe COVID-19, respectively ($p < 0.05$, Table S4).

Next, the BCV of each candidate protein was determined, and cumulative progression rates to severe COVID-19 were compared between the patients with high and low serum protein levels (Figure 4). Patients in the high-IGLV3-19, high-CRP, and high-LBP groups showed significantly higher progression rates, respectively, while patients in the low-BNC2, low-RBP4, low-TTR, low-APOC1, low-APOA2, low-KNG1, low-PF4, and low-SERPINA4 groups showed significantly higher progression rates (log-rank $p < 0.05$).

3.4 | Proteome dynamics during disease progression to severe COVID-19

To further investigate the dynamics of DEPs in patients with disease progression to severe COVID-19, we examined longitudinal changes in serum DEPs sampled at sequential time points (median, 7 time points;

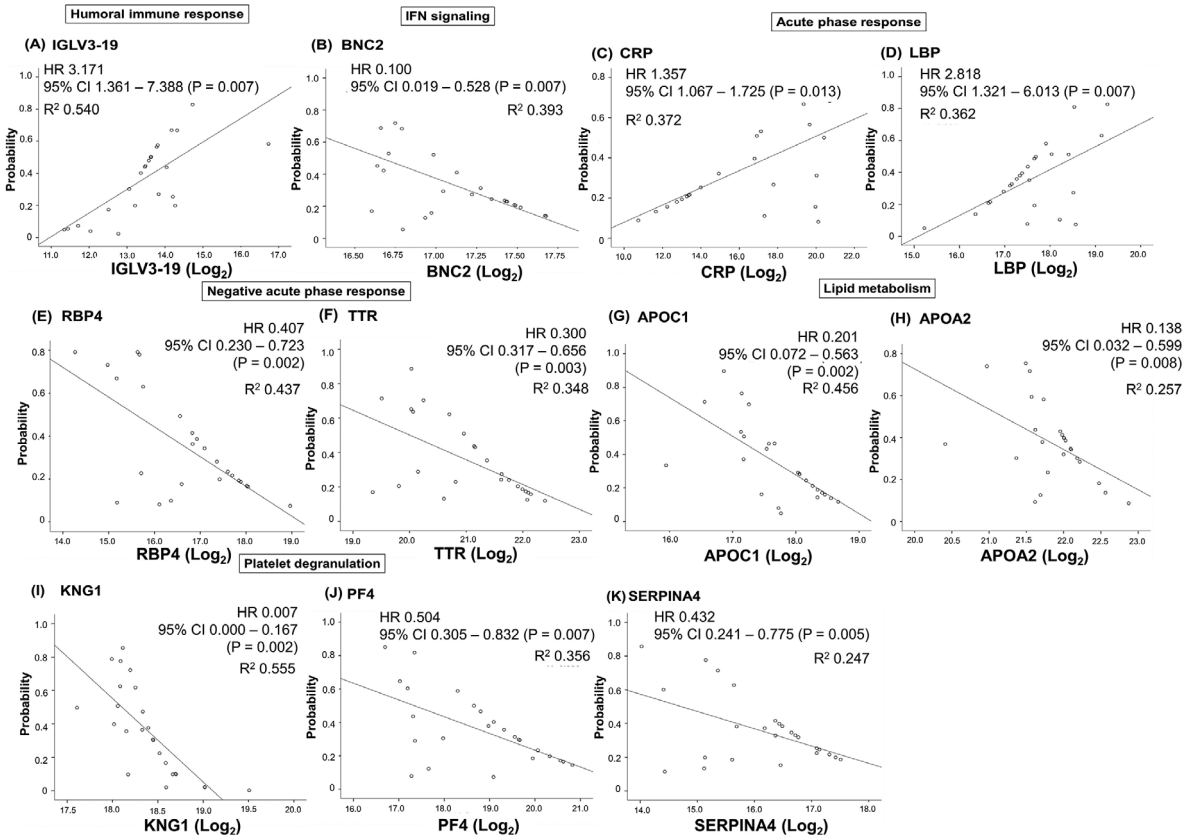

FIGURE 3 Estimated probability of disease progression to severe COVID-19, according to serum level of each protein. The risk probability was estimated using a Cox proportional hazard regression model. Data sampled at the time of admission for both severe and non-severe groups were assessed. HR, hazard ratio; CI, confidence interval.

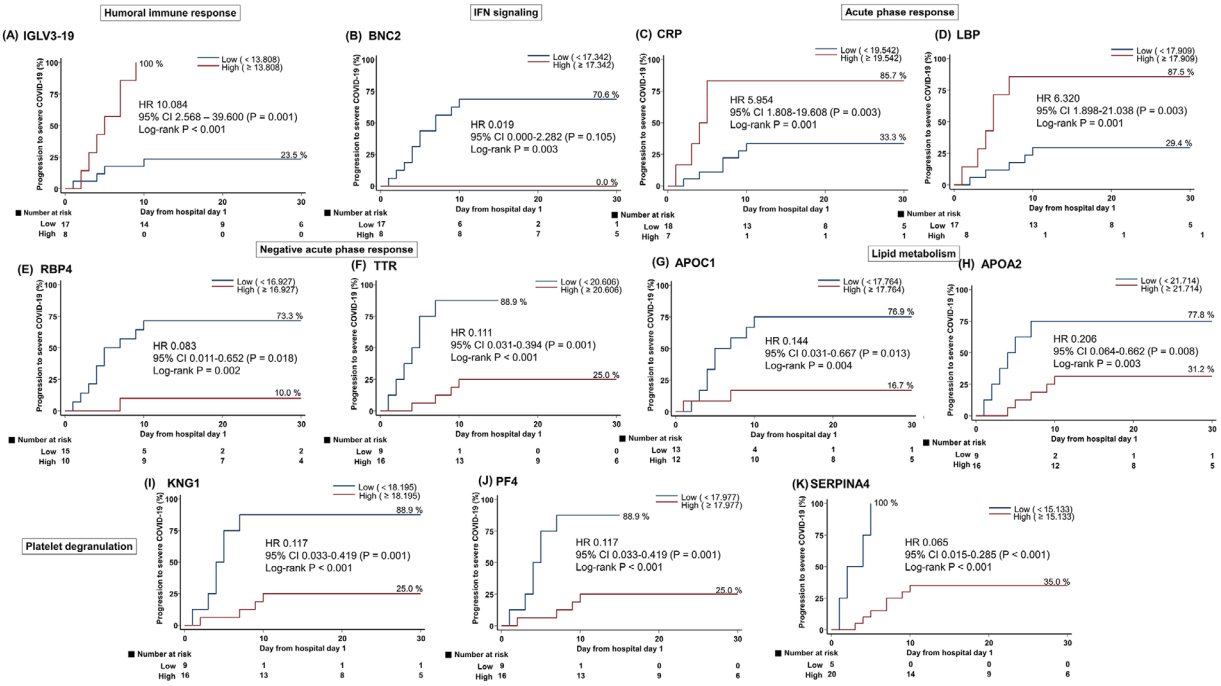

FIGURE 4 Kaplan-Meier estimates of cumulative clinical progression to severe COVID-19. Cumulative progression estimates to severe COVID-19 are present in patients with high serum levels (coloured in blue) and in patients with low serum levels (coloured in red). The best cut-off value of each protein was determined using maximally selected log-rank statistics. Data sampled at the time of admission for both severe and non-severe groups were assessed. HR, hazard ratio; CI, confidence interval.

TABLE 3 Linear mixed effect model to identify which longitudinal changes of candidate protein markers were predictive of disease progression

Proteins	Non-severe COVID-19			Severe COVID-19			Non-severe vs. severe
	B	SE	*p* value	B	SE	*p* value	FDR adjusted *p* value
IGLC2	−0.003	0.004	0.522	0.077	0.033	0.023	0.026
LEFTY2	−0.001	0.002	0.793	−0.086	0.017	0.000	<0.001
BNC2	0.001	0.003	0.801	−0.047	0.020	0.017	0.023
HGFAC	0.003	0.003	0.253	−0.081	0.022	0.000	<0.001
SEPP1	−0.002	0.002	0.379	−0.041	0.017	0.019	0.033
GC	0.000	0.002	0.963	−0.064	0.016	0.000	<0.001
APOC1	0.002	0.003	0.513	−0.068	0.023	0.004	0.006
APOM	0.002	0.002	0.331	−0.051	0.016	0.002	0.003
F9	0.003	0.003	0.275	0.103	0.021	0.000	<0.001
F10	−0.004	0.002	0.008	0.062	0.012	0.000	<0.001
SERPINA1	−0.009	0.003	0.004	0.099	0.023	0.000	<0.001
SERPINA3	−0.007	0.004	0.051	0.175	0.027	0.000	<0.001
AGT	−0.008	0.003	0.004	0.104	0.020	0.000	<0.001
A2M	0.009	0.002	0.000	−0.056	0.014	0.000	<0.001
KNG1	−0.005	0.002	0.004	−0.036	0.013	0.006	0.026
APOA1	0.001	0.002	0.653	−0.093	0.017	0.000	<0.001
CRP	−0.027	0.015	0.064	0.254	0.108	0.021	0.016
RBP4	0.011	0.004	0.011	−0.091	0.032	0.005	0.003
AHSG	0.004	0.003	0.156	−0.136	0.022	0.000	<0.001
TTR	0.011	0.004	0.011	−0.081	0.032	0.013	0.008
ALB	0.005	0.003	0.133	−0.129	0.026	0.000	<0.001
TF	0.003	0.003	0.237	−0.110	0.019	0.000	<0.001
HRG	0.012	0.004	0.002	−0.072	0.028	0.011	0.005
SERPING1	0.000	0.002	0.821	0.066	0.016	0.000	<0.001
APOA4	−0.002	0.004	0.594	−0.082	0.033	0.014	0.025
SAA1	−0.025	0.016	0.106	0.243	0.116	0.038	0.030
SAA2	−0.035	0.017	0.035	0.230	0.123	0.065	0.041
LBP	−0.017	0.004	0.000	0.059	0.033	0.076	0.030
ITIH2	0.002	0.002	0.424	−0.076	0.014	0.000	<0.001
PON1	0.006	0.002	0.011	−0.082	0.017	0.000	<0.001
TKT	−0.005	0.006	0.399	0.151	0.046	0.001	0.002
SERPINA4	0.010	0.004	0.022	−0.084	0.032	0.009	0.006
MAN1A1	−0.008	0.006	0.133	0.145	0.042	0.001	0.001
IGFALS	−0.001	0.003	0.606	−0.087	0.019	0.000	<0.001
ITIH4	−0.007	0.008	0.404	0.251	0.060	0.000	<0.001
PIK3C2B	−0.007	0.007	0.291	0.105	0.050	0.038	0.034
L1TD1	0.005	0.002	0.026	−0.080	0.017	0.000	<0.001
PRG4	−0.003	0.005	0.614	0.156	0.041	0.000	<0.001
MAN1C1	0.049	0.025	0.050	0.688	0.184	0.000	0.001
APOA2	0.002	0.003	0.547	−0.115	0.020	0.000	<0.001
IGLV3-19	0.007	0.008	0.364	0.033	0.058	0.568	0.695
IGHA2	−0.006	0.002	0.000	0.011	0.012	0.366	0.178
PF4	−0.012	0.006	0.054	−0.096	0.045	0.036	0.077

(Continues)

TABLE 3 (Continued)

Proteins	Non-severe COVID-19			Severe COVID-19			Non-severe vs. severe
	B	SE	*p* value	B	SE	*p* value	FDR adjusted *p* value
ALDOA	−0.003	0.008	0.727	0.022	0.059	0.713	0.707
HSP90AA1	0.009	0.007	0.196	0.079	0.053	0.142	0.211
PI16	0.019	0.010	0.045	0.001	0.071	0.994	0.808

Abbreviations: B, β-estimate; SE, standard error; FDR, false discovery rate.

IQR, 6–11 time points) from each patient during hospitalization (Figure S1, Table S1). The longitudinal changes in 40 DEPs were significantly different between non-severe patients and patients who were progressing to severe COVID-19 (adjusted $p < 0.05$; Table 3, Figure S5). The abundances of CRP, SAA1, SAA2, SERPINA3, ITIH4, FIX, FX, SERPINA1, SERPING1, IGLC2, AGT, TKT, MAN1A1, PIK3C2B, PRG4, and MAN1C1 increased as disease progressed ($p < 0.05$); the level of LBP tended to increase. The abundance of RBP4, TTR, ALB, TF, LEFTY2, AHSG, SEPP1, A2M, KNG1, HRG, SERPINA4, BNC2, APOA1, APOA2, APOA4, APOC1, APOM, PON1, HGFAC, GC, ITIH2, IGHALS, and L1TD1 persistently decreased as the disease progressed to severe COVID-19 ($p < 0.05$).

4 | DISCUSSION

Here, we performed proteome profiling using serum samples from patients with COVID-19 at admission and serially collected more during hospitalization. We aimed to identify potential early prognostic protein markers that could stratify high-risk patients prior to disease progression. We then further explored DEPs, which exhibit longitudinal changes that effectively reflect the clinical progression of the disease.

We found that IGLV3-19 and BNC2 could be potentially important prognostic factor for severe COVID-19. We observed increased abundance of IGLV3-19 at admission in patients who finally progressed to severe COVID-19; however, we did not find evidence of paired abundance in other immunoglobulin segments (i.e., IGHV families) in this study. Further studies of serum from larger cohorts of patients are required to more precisely understand immunoglobulin expression in severe COVID-19. Nevertheless, the increased abundance of IGLV3-19 in severe COVID-19 is a characteristic immune response that can be detected in advance of disease progression. This finding may help direct the development of vaccines as well as therapeutic strategies [22]. BNC2 might play a role in the upregulation of interferon-stimulated genes (ISGs) in the IFN signalling pathway [21]. There is a reportedly impaired IFN type 1 response with low levels of IFN-1 and ISGs associated with persistent blood viral load in severe COVID-19 patients [23, 24]. We speculate that dysregulation of BNC2 might affect IFN type 1 response and reflect a high risk for progression to severe COVID-19.

Of interest, which has not been reported yet, MAN1A1 and MAN1C1 persistently increased in the severe group subjects, and was significantly associated with the risk of disease progression. These are Golgi alpha-1,2-mannosidases, which are primarily involved in N-glycosylation that covalently adds glycan residues to the asparagine of proteins. This post-translational modification is increased in certain pathological conditions [25]. We speculate that MAN1A1 and MAN1C1 may reflect glycoproteomic changes in the host immune process in a patient with SARS-CoV-2 infection, which allows for stratification of the severity of inflammatory reactions in COVID-19 patients. Understanding significant changes in these proteins might provide insights for therapeutic strategies in COVID-19. Inhibited N-glycosylation of angiotensin I-converting enzyme 2 (ACE2) reportedly disrupts membrane fusion between SARS-CoV and host endosomes [26, 27]. In this regard, patients infected with SARS-CoV-2 might benefit from the use of inhibitors of N-glycosylation (e.g., α-glucosidase inhibitors), although further experimental studies are needed to confirm this.

Shen et al. reported that proteins, including multiple lipoproteins (APOC1, APOM), acute phase proteins (SAA1, SAA2, SAA4, CRP, SERPINA3), and proteins involved in platelet degranulation (PPBP, PF4), differentiate between severe and non-severe patients. Messner et al. reported 27 proteins involved in inflammation including IL-6 signalling (TNF, IFN-γ, ITIH4, HP, LRG1, LGALS3BP), coagulation (fibrinogen, coagulation factors, SERPINA10), acute phase response (CRP, LBP, ALB, SAA1, SAA2), and complement pathway (C1R, C1S, C8A, CFB, CFI, CFH) are differentially expressed according to disease severity. In this study, we verified that a number of proteins belonging to the acute phase response, lipid metabolism, and platelet degranulation were differently expressed between the groups of non-severe and severe patients [11, 12]. In contrast, we could not verify differential abundance of proteins belonging to the complement system. We further identified potential biomarkers differentially expressed preceding clinical deterioration in severe group and described longitudinal dynamics of the potential biomarkers during COVID-19 disease progression.

Acute phase proteins such as CRP and LBP initiate the host innate immune response to viral infection. LBP transfers lipopolysaccharides (LPS) to CD14, which subsequently bind to the toll-like receptor 4 (TLR4)/myeloid differentiation factor 2 (MD-2) complex. The CD14-LPS-MD2-TLR4 complex initiates the signalling cascade of pro-inflammatory cytokines such as IL-6, IL-8, and TNF-a [28]. The interactions of LBP and pro-inflammatory cytokines in viral infections have been previously documented [12, 29]. The hepatic synthesis of RBP4 and TTR generally decreases during acute inflammatory states, and

although the underlying pathomechanisms still remain to be clarified, a reduction in serum RBP4 and TTR has been reported as a predictor of poor clinical outcome in other diseases such as sepsis and critical ill patients [30, 31].

Dysregulation of multiple apolipoproteins (e.g., APOA1, APOA2, APOC1, and APOM) and suppressed platelet degranulation in the serum of patients with severe COVID-19 have been previously reported [11, 12, 32], which is in accordance with the findings of the present study. Currently, it is unclear whether SARS-CoV-2 infection and the pathogenesis of severe COVID-19 directly alter platelet activity [33]. However, considering that platelets directly recognize pathogens and release antimicrobial mediators by degranulation in viral infections, dysregulation of platelet degranulation in SARS-CoV-2 infection implies a possible impaired immune response in severe COVID-19 patients [34]. Our proteomic analysis further elucidates the predictive role of initial and serial measurement of these protein abundances over clinical outcomes in patients with COVID-19.

One of the limitations of our study is that our findings have not yet been externally validated. However, we performed proteomic analysis of sequential time points from each patient, which enabled us to identify reliable DEPs that were consistently differentially expressed in severe disease status, at initial assessment, and during hospitalization. Our results should be validated in an independent patient cohort. In addition, the number of included patients was small ($n = 25$). Future larger prospective studies are warranted to validate these potential biomarkers.

In summary, this study investigated the proteomics dynamics through longitudinal proteomic profiling on serum samples from patients with COVID-19, and demonstrates proteins significantly changed in severe versus non-severe disease. Our findings may give some new insights into host response to SARS-CoV-2 infection, which are valuable for understanding of disease progression to severe COVID-19. Our data could be integrated with additional data in the near future to develop protein biomarkers, which might enable better predicting and monitoring progression to severe COVID-19.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in ProteomeXchange Consortium via the PRIDE at https://www.ebi.ac.uk/pride/archive, reference number PXD023686.

ORCID

Jee-Soo Lee https://orcid.org/0000-0002-6633-4631
Dohyun Han https://orcid.org/0000-0002-0841-1598

REFERENCES

1. Weiss, P., & Murdoch, D. R. (2020). Clinical course and mortality risk of severe COVID-19. *Lancet*, *395*, 1014–1015.
2. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, *395*, 497–506.
3. Wu, S., Du, Z., Shen, S., Zhang, B., Yang, H., Li, X., Cui, W., Cheng, F., & Huang, J. (2020). Identification and validation of a novel clinical signature to predict the prognosis in confirmed COVID-19 patients. *Clinical Infectious Diseases*, *71*(12), 3154–3162.
4. Khartabil, T. A., Russcher, H., van der Ven, A., & de Rijke, Y. B. (2020). A summary of the diagnostic and prognostic value of hemocytometry markers in COVID-19 patients. *Critical Reviews in Clinical Laboratory Sciences*, *57*, 415–431.
5. Li, Q., Cao, Y., Chen, L., Wu, D., Yu, J., Wang, H., He, W., Chen, L., Dong, F., Chen, W., Chen, W., Li, L., Ran, Q., Liu, Q., Ren, W., Gao, F., Chen, Z., Gale, R. P., & Hu, Y. (2020). Hematological features of persons with COVID-19. *Leukemia*, *34*, 2163–2172.
6. Tan, L., Wang, Q., Zhang, D., Ding, J., Huang, Q., Tang, Y.-Q., Wang, Q., & Miao, H. (2020). Lymphopenia predicts disease severity of COVID-19: A descriptive and predictive study. *Signal Transduction and Targeted Therapy*, *5*, 33.
7. Whetton, A. D., Preston, G. W., Abubeker, S., & Geifman, N. (2020). Proteomics and informatics for understanding phases and identifying biomarkers in COVID-19 disease. *Journal of Proteome Research*, *19*, 4219–4232.
8. Mehta, P., Mcauley, D. F., Brown, M., Sanchez, E., Tattersall, R. S., & Manson, J. J. (2020). COVID-19: Consider cytokine storm syndromes and immunosuppression. *Lancet*, *395*, 1033–1034.
9. Wendt, R., Kalbitz, S., Lübbert, C., Kellner, N., Macholz, M., Schroth, S., Ermisch, J., Latosisnka, A., Arnold, B., Mischak, H., Hartmut Beige, J., & Metzger, J. (2020). Urinary proteomics associates with COVID-19 severity: Pilot proof-of-principle data and design of a multicentric diagnostic study. *Proteomics*, 2000202.
10. Jankowski, J. (2020). Proteomic biomarkers to guide stratification for Covid-19 treatment: Exemplifying a path forward toward implementation? *Proteomics*, 2000229.
11. 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., … Guo, T. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell*, *182*, 59-72.e15.
12. Messner, C. B., Demichev, V., Wendisch, D., Michalick, L., *et al*. (2020). Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Systems*, *11*, 11–24 e14.
13. Almeida, A., Faustino, M. A. F., & Neves, M. (2020). Antimicrobial photodynamic therapy in the control of COVID-19. *Antibiotics (Basel)*, *9*.
14. Rhee, S. J., Han, D., Lee, Y., Kim, H., Lee, J., Lee, K., Shin, H., Kim, H., Lee, T. Y., Kim, M., Kim, S. H., Ahn, Y. M., Kwon, J. S., & Ha, K. (2020). Comparison of serum protein profiles between major depressive disorder and bipolar disorder. *BMC Psychiatry [Electronic Resource]*, *20*, 145.
15. Jeon, H. S., Lee, S. M., Jung, Y. M., Oh, S., Park, J. K., Lee, E. B., Park, C.-W., Park, J. S., Han, D., & Jun, J. K. (2020). Proteomic biomarkers in mid-trimester amniotic fluid associated with adverse pregnancy outcomes in patients with systemic lupus erythematosus. *Plos One*, *15*, e0235838.
16. Kim, Y.-S., Han, D., Kim, J., Kim, D. W., Kim, Y.-M., Mo, J.-H., Choi, H.-G., Park, J.-W., & Shin, H.-W. (2019). In-depth, proteomic analysis of nasal secretions from patients with chronic rhinosinusitis and nasal polyps. *Allergy, Asthma & Immunology Research*, *11*, 691–708.
17. Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M. O., & Aebersold, R. (2011). mProphet: Automated data processing and statistical validation for large-scale SRM experiments. *Nature Methods*, *8*, 430–435.
18. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yılmaz, Ş., Tiwary, S., Cox, J., Audain, E., Walzer, M., … Vizcaíno, J. A. (2019). The PRIDE

database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res*, *47*, D442–D450.
19. Hothorn, T., & Lausen, B. (2003). On the exact distribution of maximally selected rank statistics. *Computational Statistics & Data Analysis*, *43*, 121–137.
20. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.
21. Urgard, E., Reigo, A., Reinmaa, E., Rebane, A., & Metspalu, A. (2017). Human basonuclin 2 up-regulates a cascade set of interferon-stimulated genes with anti-cancerous properties in a lung cancer model. *Cancer Cell International*, *17*, 18.
22. Zohar, T., & Alter, G. (2020). Dissecting antibody-mediated protection against SARS-CoV-2. *Nature Reviews Immunology*, *20*, 392–394.
23. Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., Breillat, P., Carlier, N., Gauzit, R., Morbieu, C., Pène, F., Marin, N., Roche, N., Szwebel, T.-A., Merkling, S. H., …, Terrier, B. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*, *369*, 718–724.
24. Lee, J. S., & Shin, E.-C. (2020). The type I interferon response in COVID-19: Implications for treatment. *Nature Reviews Immunology*, 585–586.
25. Joenvaara, S., Saraswat, M., Kuusela, P., Saraswat, S., Agarwal, R., Kaartinen, J., Järvinen, A., & Renkonen, R. (2018). Quantitative N-glycoproteomics reveals altered glycosylation levels of various plasma proteins in bloodstream infected patients. *Plos One*, *13*, e0195006.
26. Williams, S. J., & Goddard-Borger, E. D. (2020). alpha-Glucosidase inhibitors as host-directed antiviral agents with potential for the treatment of COVID-19. *Biochemical Society Transactions*, *48*, 1287–1295.
27. Zhao, X., Guo, F., Comunale, M. A., Mehta, A., Sehgal, M., Jain, P., Cuconati, A., Lin, H., Block, T. M., Chang, J., & Guo, J.-T. (2015). Inhibition of endoplasmic reticulum-resident glucosidases impairs severe acute respiratory syndrome coronavirus and human coronavirus NL63 spike protein-mediated entry by altering the glycan processing of angiotensin I-converting enzyme 2. *Antimicrobial Agents and Chemotherapy*, *59*, 206–216.
28. Ding, P.-H., & Jin, L. J. (2014). The role of lipopolysaccharide-binding protein in innate immunity: a revisit and its relevance to oral/periodontal health. *J Periodontal Res*, *49*, 1–9.
29. Van Gucht, S., Van Reeth, K., Nauwynck, H., & Pensaert, M. (2005). Porcine reproductive and respiratory syndrome virus infection increases CD14 expression and lipopolysaccharide-binding protein in the lungs of pigs. *Viral Immunology*, *18*, 116–126.
30. Koch, A., Weiskirchen, R., Sanson, E., Zimmermann, H. W., Voigt, S., Dückers, H., Trautwein, C., & Tacke, F. (2010). Circulating retinol binding protein 4 in critically ill patients before specific treatment: Prognostic impact and correlation with organ function, metabolism and inflammation. *Critical Care*, *14*, R179.
31. Devakonda, A., George, L., Raoof, S., Esan, A., Saleh, A., & Bernstein, L. H. (2008). Transthyretin as a marker to predict outcome in critically ill patients. *Clinical Biochemistry*, *41*, 1126–1130.
32. Nie, S., Zhao, X., Zhao, K., Zhang, Z., *et al*. (2020). Metabolic disturbances and inflammatory dysfunction predict severity of coronavirus disease 2019 (COVID-19): a retrospective study. medRxiv 2020; https://doi.org/10.1101/2020.03.24.20042283.
33. Manne, B. K., Denorme, F., Middleton, E. A., Portier, I., Rowley, J. W., Stubben, C., Petrey, A. C., Tolley, N. D., Guo, L., Cody, M., Weyrich, A. S., Yost, C. C., Rondina, M. T., & Campbell, R. A. (2020). Platelet gene expression and function in patients with COVID-19. *Blood* 1317–1329.
34. Assinger, A. (2014). Platelets and infection - An emerging role of platelets in viral infection. *Frontiers in Immunology*, *5*, 649.

SUPPORTING INFORMATION

Additional supporting information may be found online https://doi.org/10.1002/pmic.202000278 in the Supporting Information section at the end of the article.

How to cite this article: Lee, J.-S., Han, D., Kim, S. Y., Hong, K. H., Jang, M.-j., Kim, M. J., Kim, Y.-g., Park, J. H., Cho, S. I., Park, W. B., Lee, K. B., Shin, H. S., Oh, H. S., Kim, T. S., Park, S. S., & Seong, M.-W. (2021). Longitudinal proteomic profiling provides insights into host response and proteome dynamics in COVID-19 progression. *Proteomics*, *21*, e2000278. https://doi.org/10.1002/pmic.202000278