

# Identification of Serum Prognostic Biomarkers of Severe COVID-19 by Quantitative Proteomic Approach

**Yayoi Kimura**

Advanced Medical Research Center, Yokohama City University

**Yusuke Nakai**

Advanced Medical Research Center, Yokohama City University

**Jihey Shin**

Advanced Medical Research Center, Yokohama City University

**Miyui Hara**

School of Medicine Medical Course Biostatistics, Yokohama City University

**Yuriko Takeda**

School of Medicine Medical Course Biostatistics, Yokohama City University

**Sousuke Kubo**

Department of Pulmonology, Yokohama City University School of Medicine

**Sundararaj S Jeremiah**

Department of Microbiology, Yokohama City University School of Medicine

**Yoko Ino**

Advanced Medical Research Center, Yokohama City University

**Tomoko Akiyama**

Advanced Medical Research Center, Yokohama City University

**Kayano Moriyama**

Advanced Medical Research Center, Yokohama City University

**Kazuya Sakai**

School of Medicine Medical Course Emergency Medicine, Yokohama City University

**Ryo Saji**

School of Medicine Medical Course Emergency Medicine, Yokohama City University

**Mototsugu Nishii**

School of Medicine Medical Course Emergency Medicine, Yokohama City University

**Hideya Kitamura**

Kanagawa Cardiovascular and Respiratory Center

**Kota Murohashi**

Kanagawa Cardiovascular and Respiratory Center

**Kouji Yamamoto**

School of Medicine Medical Course Biostatistics, Yokohama City University

**Takeshi Kaneko**

Department of Pulmonology, Yokohama City University School of Medicine

**Ichiro Takeuchi**

School of Medicine Medical Course Emergency Medicine, Yokohama City University

**Eri Hagiwara**

Kanagawa Cardiovascular and Respiratory Center

**Takashi Ogura**

Kanagawa Cardiovascular and Respiratory Center

**Hideki Hasegawa**

Influenza Research Center, National Institute of Infectious Diseases

**Tomohiko Tamura**

Department of Immunology, Yokohama City University School of Medicine

**Takeharu Yamanaka**

School of Medicine Medical Course Biostatistics, Yokohama City University

**Akihide Ryo** (✉ [aryo@yokohama-cu.ac.jp](mailto:aryo@yokohama-cu.ac.jp))

Department of Microbiology, Yokohama City University School of Medicine

---

**Research Article**

**Keywords:** COVID-19, Proteomics, Biomarker, Prognosis, Diagnosis

**Posted Date:** February 10th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-156575/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The COVID-19 pandemic is an unprecedented threat to humanity provoking global health concerns. Since the etio-pathogenesis of this illness is not fully characterized, the prognostic factors enabling treatment decisions have not been well documented. An accurate prediction of the disease progression can aid in appropriate patient categorization to determine the best treatment option. Here, we have introduced a proteomic approach utilizing data-independent acquisition mass spectrometry (DIA-MS) to identify the serum proteins closely associated with the prognosis of COVID-19. We observed 27 proteins to be differentially expressed between the cohorts of severely ill COVID-19 patients with adverse and favorable prognosis. Ingenuity pathway analysis revealed that 15 out of the 27 proteins might be regulated by cytokine signalling relevant to interleukin (IL)-1b, IL-6 and tumor necrosis factor (TNF), and their differential expression was possibly implicated in the systemic inflammatory response and cardiovascular disorders. We further evaluated the practical prognosticators for the clinical prognosis of severe COVID-19 patients. Subsequent ELISA analyses further uncovered that CHI3L1 and IGFALS could be potent prognostic markers with a high sensitivity. Our findings can help in formulating a diagnostic approach for accurately discriminating severe COVID-19 patients and provide appropriate treatment based on their predicted prognosis.

## Introduction

Coronavirus disease 2019 (COVID-19) is a highly transmittable respiratory infection caused by the novel positive-sense, single-stranded RNA virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that emerged in Wuhan, China in 2019. Despite containment efforts, due to rapid person-to-person transmission the disease disseminated into a pandemic and is still spreading<sup>1</sup>. The molecular mechanisms behind disease progression leading to respiratory distress in COVID-19 patients are still unknown and no effective anti-viral therapies for COVID-19 have been established till date<sup>2</sup>. In order to optimize allocations of limited health care resources to the neediest patients, it is crucial to accurately predict the progress and prognosis of patients with this disease. In addition, severity risk management for patients contributes to further reduction in mortality<sup>3,4</sup>.

Most COVID-19 patients exhibit mild symptoms without dyspnea and abnormal chest imaging or moderate respiratory symptoms with pneumonia. They usually recover from illness with or without supportive treatment. About 20% of COVID-19 patients develop respiratory distress and requires immediate oxygen supplementation. A subset of these patients become critically ill developing rapid respiratory failure and severe hypoxemia, and requires immediate intensive care to prevent fatality. Considering the diverse variations in clinical manifestations of COVID-19, stratifying patients who are at risk for developing severe disease with adverse prognosis is crucial for selecting appropriate treatment strategies. For this purpose, identifying novel biological indicators that can serve as precise prognostic biomarkers is necessary to help clinicians make better clinical decisions and provide appropriate therapeutic strategies in earlier stages. To date, several clinical and biochemical parameters are used for predicting the severity of COVID-19 such as; C-reactive protein (CRP), serum amyloid A (SAA), interleukin

(IL)-6, lactate dehydrogenase (LDH), white blood cell count, D-dimer, cardiac troponin and platelet count<sup>3,5</sup>. In addition, multiple serological factors involved in the severity of COVID-19 have been identified by proteomic approach using serum of the patients<sup>6-10</sup>. Most of these studies observed the protein profiles involved in systemic and/or local inflammation, and accompanying organ damage or dysfunction. Although the currently available serological biomarkers can predict severe disease, the markers that predict clinical prognosis and mortality of severe COVID-19 patients are not yet reported.

In this study, we utilized the recently developed mass spectrometry technology with the data independent acquisition (DIA-MS) approach to identify the serum proteins closely associated with the disease prognosis. Using ELISA analyses, we further delineate the practical prognosticators for the clinical prognosis of severe COVID-19 patients. Consequently, we identified putative biomarkers that can indicate disease progression and adverse prognosis of the patients. These biomarkers shed light on a novel diagnostic approach that can serve to segregate COVID-19 patients based on the expected clinical prognosis and channelize appropriate measures for their management.

## Results

### Identification of serum proteins associated with favorable or adverse outcomes of severe COVID-19 patients by quantitative proteomic analysis

To identify the serological biomarkers involved in the favorable/adverse prognosis of severe COVID-19 patients, we performed a comparative proteomic analysis with DIA-MS (Fig. 1A). In the discovery study, we obtained the MS data from serum samples collected within a few days after the start of special inpatient intervention of 10 severe COVID-19 patients with different prognosis (five adverse and five favorable). By utilizing our customized spectral DIA library containing information of 1,534 human serum proteins, we identified 656 proteins to be differentially expressed in sera at the protein false discovery rate (FDR) 1% level. Among them, 495 proteins were selected for further statistical analysis with Perseus (Table S1). Subsequent principal component analysis (PCA) score plot visualized the distribution of the samples and revealed obvious separation trend between the two groups (Fig. 1B). To identify proteins that changed prominently along the disease prognosis, the significant changes of proteins in the severe ill patients with adverse prognosis were next analyzed using volcano plots. Consequently, 16 up-regulated proteins and 11 down-regulated proteins were selected being associated with the adverse prognosis of COVID-19 with statistical significance ( $p$ -value < 0.01, fold change (difference) > 2) (Fig. 1C and Table 1). Indeed, the heat map analysis exhibited hierarchical clustering of these proteins based on the expression levels be correlated with disease prognosis of severe COVID-19 patients (Fig. 1D). To investigate the biological processes affecting the severity of COVID-19, we carried out with an upstream analysis within the framework of the Ingenuity Pathway Analysis (IPA). The results showed that several increasing and decreasing proteins in the serum of the severe patients with adverse prognosis might be regulated by proinflammatory cytokines (Table S2). Notably, out of 27 differentially expressed proteins, 15 proteins were found to be regulated by IL-1 $\beta$ , IL-6 and tumor necrosis factor (TNF), which are seen at markedly higher levels in most of severe COVID-19 patients<sup>11,12</sup> (Fig. 2). In addition, a disease and disorders

enrichment analysis suggested that the several differentially expressed proteins could be associated with cardiovascular disorders (Table S3). The result was consistent with the hypothesis that COVID-19 causes cardiovascular diseases, including myocardial injury and venous thromboembolism<sup>13</sup>. Simultaneously, inflammatory response, such as the degranulation of neutrophils was also enriched, as reported in the literature<sup>7-10</sup>. Furthermore, most of proteins among the 27 differentially expressed ones demonstrated to form an interconnected network as revealed by the STRING database (Figure S1).

### **Identification of putative biomarkers for predicting prognosis of critical COVID-19**

In order to search for the practical prognostic indicators for severe COVID-19 patients, we focused on two proteins, namely CHI3L1, and insulin-like growth factor binding protein acid labile subunit (IGFALS), which were most significantly correlated positively or inversely with the adverse prognosis of the severe COVID-19 patients in statistical analyses (Table 1). We excluded myoglobin (MB) since it has previously been reported as a prognostic marker<sup>9</sup>. To evaluate the candidates for clinical benefit, the relative expression changes associated with the prognosis of 41 severe COVID-19 patients (10 adverse and 31 favorable; Table S4), we analyzed the serum levels of these proteins using ELISA assays. The clinical information of the recruited patients enrolled in the evaluation study is presented in Table S4. These parameters, excluding ECMO care and outcome of death, were not different among adverse and favorable groups. On the other hand, ELISA assay for the levels of these proteins in severe COVID-19 patients with different prognosis showed significant changes between the adverse and favorable prognosis groups ( $p < 0.005$ ), suggesting that the serum levels of these proteins also correlate with the adverse prognosis of the severe COVID-19 patients (Fig. 3A). We further assessed the predictive ability of serum levels of CHI3L1, and IGFALS, to detect the COVID-19 patients with adverse prognosis using the receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) [95% confidence intervals (CI)] of a model using each value of CHI3L1, or IGFALS in the present study set was 0.852 [0.710–0.958], or 0.871 [0.755–0.968], respectively (Fig. 3B). Internal validation of the final models for CHI3L1 and IGFALS was performed by bootstrap resampling, where they maintained high accuracy (optimism-adjusted AUC = 0.853, and 0.872, respectively). These values were shown to be more reliable than those of pre-existing biomarkers, CRP (0.49 [0.258–0.716]) or D-dimer (0.385 [0.202–0.582]) (Figure S2). More precise statistical analysis in AUC also corroborated that CHI3L1 and IGFALS were markers with higher reliability, compared to CRP or D-dimer (Fig. 3B). Furthermore, the AUC [95% CI] of a model using both values of CHI3L1 and IGFALS was 0.91 [0.797–0.990]. Consequently, the serum expression levels of these two proteins can enhance the clinical diagnostics by providing a precise indication of the outcome of a severely ill COVID-19 patient.

## **Discussion**

COVID-19 displays variable illness or symptoms ranged from asymptomatic cases leading to the spontaneous recovery, to acute respiratory distress syndrome (ARDS) characterized by respiratory failure and diffuse alveolar damage<sup>14</sup>. While most patients with severe respiratory disorders recover from the illness, a substantial number of people die of respiratory failure and/or systemic complications. Choosing

the right target population which is most susceptible to adverse prognosis would be the right way for channelizing intensive medical management. Identifying putative risk factors and/or biomarkers for predicting prognosis of critical COVID-19 is a felt need which might help clinicians and health care professionals to select or devise an appropriate management protocol for COVID-19 patients. For this purpose, we performed DIA-MS-based proteomic analysis, which has the potential to discover proteins related to adverse prognosis even without prior knowledge, using the serum of COVID-19 patients. Consequently, we identified 27 candidate proteins that were selectively increased or decreased along with the adverse prognosis. Subsequent statistical analysis with ROC curve explored two putative prognostic indicators that can be useful for predicting the most likely clinical prognosis of severe COVID-19 patients. As a result, we identified two putative newly-developed prognostic indicators, namely CHI3L1 and IGFALS.

Several studies have reported that most severe cases of COVID-19 exhibit marked increase in serum pro-inflammatory cytokines<sup>11,12,15</sup>. Therefore, the current understanding of the disease suggests that cytokine storm along with the immunological dysregulation triggered by the viral replication phase contribute to the progression of severe ARDS and multiple organ failures in COVID-19<sup>11,16</sup>. However, IL-6 levels in COVID-19 patients are lower than the median values typically reported in ARDS<sup>17,18</sup>, and there could be other unidentified determinants that define COVID-19 severity. In our current study, we analyzed the molecular relevance of CHI3L1 and IGFALS by an upstream analysis within the framework of the IPA. Consequently, we found that the expression levels of these proteins were regulated by pro-inflammatory cytokine such as IL-1 $\beta$ , IL6 or TNF. This finding indicates that our newly-developed biomarkers could be “surrogate” markers of the proinflammatory cytokine network and cascade.

CHI3L1 (chitinase-3-like protein 1), also termed YKL-40, is a protein that binds tightly with chitin but lacks chitinase activity. Our current study found that the serum CHI3L1 levels increased depending on the disease severity and adverse prognosis of COVID-19 patients. Parallel to this finding, previous studies have shown the correlation of high blood CHI3L1 levels with increased risk of death from various causes, including cardiovascular disease<sup>19,20</sup>. Additionally, elevated levels of the circulating CHI3L1 are associated with patients with idiopathic pulmonary fibrosis (IPF)<sup>21</sup>. Immunohistochemistry also shows that the expression levels of CHI3L1 are enhanced in bronchiolar epithelial cells and alveolar macrophages adjacent to fibrotic lesions of the patients with IPF, suggesting the possible involvement of CHI3L1 in the fibrotic process of IPF<sup>21</sup>. These findings together suggest that CHI3L1 plays an important role in the tissue remodelling of respiratory system in COVID-19 following the massive inflammation and alveolar destruction and local tissue remodelling<sup>22,23</sup>. Therefore, higher levels of CHI3L1 might be associated with the pathogenesis of COVID-19 especially in relevance to pulmonary tissue damage and repair.

Our current study also demonstrated that IGFALS levels were down-regulated with disease severity and adverse prognosis in COVID-19. In the general state, IGFALS forms a ternary complex with IGFBP3 and Insulin-like growth factor 1 (IGF-1). The binding of IGFALS/IGFBP3 with IGF-1 was shown to prevent the interaction of IGF-1 to its receptor IGF-1R, and also reduced the stability of IGF-1 to suppress its biological

function<sup>24</sup>. It was also observed that plasma levels of IGF-1 were significantly reduced in complete IGFALS deficiency mice, suggesting the accelerated reduction of half-life without any changes in their liver or renal expression<sup>25,26</sup>. Consequently, the deficiency of IGFALS proteins causes disruption of the entire IGF-1 circulating system, without affecting glucose and insulin homeostasis<sup>25</sup>. The role of IGF-1 signaling in fibrotic processes is variable depending on its spatial and stoichiometric conditions<sup>27</sup>. Irrespective of COVID-19, IGF-1 level diminishes gradually at later fibroproliferative stage representing negative correlation with the mortality of patients with ARDS<sup>28,29</sup>. Moreover, a recent study has indicated that low serum IGF-1 levels are associated with higher risk of mortality risk in COVID-19 patients<sup>30</sup>. These finding may together suggest that serum IGF-1/IGFALS levels are directly or indirectly involved in respiratory dysfunction. However, the regulatory mechanism of IGFALS and IGF-1 in COVID-19 remain elusive and further studies will be required to determine the functional role of IGF-1/IGFALS in the pathogenesis of COVID-19.

The findings of this study could enhance the diagnosis of COVID-19 patients with severe pneumonia at high risk of mortality by combining the serum levels of two proteins closely involved in the pathogenesis in COVID-19. The capability of CHI3L1 and IGFALS for discriminating COVID-19 patients with adverse prognosis from favorable prognosis was superior to that of the existing biomarkers; CRP and D-dimer.

## Limitations Of The Study

Although we have performed a comprehensive proteomic analysis in the current study, further prospective studies will be required to validate the quality of these biomarkers. This could be achieved with a multidisciplinary approach, and a multivariable statistical analysis of these prognostic biomarkers in order to increase their accuracy of detecting clinical prognosis of severe COVID-19.

## Methods

### Human Samples

Serum samples were obtained from COVID-19 patients who were hospitalized at Yokohama City University Hospital, Yokohama City University Medical Center, and National Hospital Organization Yokohama Medical Center from February to June 2020. This research plan and protocol was approved by the Clinical Ethics Committee of Yokohama City University Hospital (B2002000048). This study was also performed with the approval of the Clinical Ethics Committee in each of the medical facilities. Informed consent was obtained from all patients and/or their guardians before serum samples collection. This study was conducted in accordance with the Declaration of Helsinki. All the data was anonymized before the analyses.

All patients in the study were diagnosed as COVID-19 according to the manual for the Detection of Pathogen 2019-nCoV of the National institute of infectious diseases in Japan. Severe COVID-19 patients were divided according to the National Institutes of Health guideline. In addition, severe patients with an

outcome of death or requiring ECMO care were designated as the patients with adverse prognosis, while others were designated as the patients with favorable prognosis. All serum samples were stored at -80°C until use and then denatured by adding an equal volume of 8 M urea solution for MS analysis.

### **LC mass spectrometry**

Desalted peptides were resuspended in 0.1% formic acid and 2% ACN containing iRT peptides (Biognosys) and then analyzed using Q Exactive™ mass spectrometer coupled with an UltiMate™ 3000 HPLC system. The mass spectrometer was operated using Xcalibur software. Peptides were loaded on a trap column (100 µm × 20 mm, C18, 5 µm, 100 Å, Thermo Fisher Scientific) and subsequently were separated on a Nano HPLC capillary column (75 µm × 180 mm, C18, 3 µm, Nikkyo Technos) at a flow rate of 300 nL/minute. Solvent A was 0.1% formic acid in 2% acetonitrile (ACN), while solvent B was 0.1% formic acid in 95% ACN. Peptides were eluted using a gradient from 2% B for 0-5 minutes, 2% to 33% B for 5-120 minutes followed by 90% B for 10 minutes, and then equilibrated for 20 min at 2% B. Data were acquired using either data-dependent acquisition (DDA) or data-independent acquisition (DIA).

### **Human sera spectral library generation**

For the comprehensive serum proteome analysis, we attempted to construct the original DIA-MS system for human serum. To construct a serum spectral library, human pooled sera purchased from KOHJIN BIO, BioWest, PANBIOTECH, and SIGMA were fractionated in three ways after removal of 14 human proteins (ALB, IgG, antitrypsin, IgA, transferrin, haptoglobin, fibrinogen, alpha2-macroglobulin, alpha1-acid glycoprotein, IgM, apolipoprotein AI, apolipoprotein AII, complement C3, and transthyretin) using a Human 14 Multiple Affinity Removal System (MARS) column (Agilent Technologies) or after compression of dynamic range of protein abundance using a Proteominer beads (Bio-Rad) according to each manufacturer's instructions. First, the immunodepleted or compressed serum was fractionated using HPLC system with a C4 reversed-phase column (Vydac) and 20 fractions were independently subjected to in-solution digestion with trypsin (Promega)<sup>31</sup>. Second, the immunodepleted or compressed serum was separated into six pieces on a 5-20 % polyacrylamide gel, followed by in-gel digestion with trypsin<sup>32</sup>. Third, the immunodepleted serum was digested with trypsin, and the resulting peptides (240 µg) were separated into 24 fractions by 3100 OFFGEL Fractionator<sup>33</sup>. After desalting using a Stage Tip<sup>34</sup>, the obtained peptides were analyzed in a DDA mode. The Q-Exactive was set to positive mode in a top-20 configuration. DDA mode analytical conditions consisted of a full MS1 scan with the resolution of 70,000 with a scan range from 350 to 1,500 *m/z*, with the AGC set to 3e<sup>6</sup> (Full MS) and 1e<sup>5</sup> (MS/MS). The normalized collision energy was set to 27. Spectral library generation from 76 DDA-MS measurements dataset were performed using Spectronaut Pulsar X (Ver.12.0.2, Biognosys) by searching against the iRT fasta database (Biognosys) and a human protein sequences of UniProtKB/Swiss-Prot database (version January 28, 2019), allowing for variable N-terminal acetylation, N-terminal carbamylation, methionine oxidation and cysteines carbamidomethylation. MS1 and MS2 tolerances were set to dynamic, and two missed cleavages were allowed. Search results were filtered to a 1% peptide-level FDR using Spectronaut Pulsar X.



## Sample preparation for DIA-MS analysis

After adding 20 ng/ $\mu$ l *E. coli* b-galactosidase (b-gal) as the internal standard 14 high abundance serum proteins (ALB, IgA, IgD, IgE, IgG, IgG [light chains], IgM, alpha-1-acid glycoprotein, alpha-1-antitrypsin, alpha-2-macroglobulin, apolipoprotein A1, fibrinogen, haptoglobin, transferrin) were removed by using a High Select™ Top14 Abundant Protein Depletion Mini Spin Columns (ThermoFisher Scientific) according to each manufacturer's instructions. After centrifugal ultrafiltration using Amicon Ultra centrifugal filters, immunodepleted serum samples were dissolved in 8 M urea solution. For verification analysis of depletion reproducible efficiency, the proteins separated by SDS-PAGE proteins were transferred to PVDF membranes and then incubated with anti-b-gal antibody (diluted 1:1,000), at room temperature (data not shown). Subsequently proteins corresponding to 2 ml of immunodepleted serum were reduced with DTT (final concentration of 10mM) and alkylated with IAA (final concentration of 25mM). The protein solutions were diluted to 2 M urea in 50 mM  $\text{NH}_4\text{HCO}_3$  and then incubated with trypsin (final concentration, 15 ng/ $\mu$ l) at 37 °C for 16 h. To prepare peptides for MS analysis, the resultant peptides were desalted by using a Stage Tip<sup>34</sup>, and subsequently eluted peptides were completely lyophilized, and kept at - 80 °C until use.

## DIA-MS analysis and data analysis

To determine protein abundance, the serum peptide samples were analyzed twice each in a DIA mode. DIA mode analytical conditions consisted of a full MS1 scan with the resolution of 70,000 full width at half-maximum (FWHM) with a scan range from 380 to 1240 m/z, with the AGC set to  $3e^6$ , followed by 40 DIA windows acquired at a resolution of 35,000 FWHM, with the AGC value to  $3e^6$ . The normalized collision energy was set to 28. DIA-MS data were analyzed using Spectronaut Pulsar X against the spectral library to identify and quantify peptides and proteins. Retention time calibration was set to the iRT peptides. The Biognosys default settings were applied for identification: excluding duplicate assay and estimation of FDRs using q-value as 0.01 for both precursors and proteins. Interference correction was activated and a minimum of 3 fragment ions and 2 precursor ions were kept for the quantitation. The area of extracted ion chromatogram (XIC) at MS/MS level were used for quantitation. Peptide quantity was measured by the mean of 1-10 best precursors, and protein quantity was calculated accordingly by summing 1-10 best peptides. Global normalization strategy and q-value sparse selection were used for cross run normalization. All other settings were set to default. All proteomics data are deposited in the ProteomeXchange Consortium via the jPOST partner repository (Project ID: PXD021702) (preview URL for reviewers: <https://repository.jpostdb.org/preview/116104560600aa7ee9ab41>, Access key: 2950). To perform downstream statistical quantitative analysis, we used Perseus (Max-Planck-Institute of Biochemistry), which is a software for functional analysis of large-scale quantitative data<sup>35</sup>. Distinct samples were categorized in the respective groups, the intensity values were  $\log_2$ -transformed and only proteins quantified in at least 70% of samples for each group were used for further analysis. Normalization was performed by width adjustment previously to the imputation of the missing values (downshift=1.8 and width=0.3). PCA score plot and volcano plot were performed with Perseus. Protein

interaction analysis was carried out with an online tool STRING (default setting)<sup>36</sup>. IPA was used for the biological analysis. According to manufacturer instructions, ELISA assays were performed to measure the circulating levels of CHI3L1 (cat# CY8088V2, MBL) and IGFALS (cat# 445907, BIOLEGEND). ROC curve analysis was performed to assess the predictive performance of CHI3LI, IGFALS, D-dimer and CRP. The optimal cut-off value was determined by Youden index. The internal validation was performed by bootstrapping and was done with 150 simulations to obtain a bootstrapped AUC.

Statistical analysis was performed using GraphPad Prism 8 software, EZR (Easy R) software (version 4.7-1), or statistical software R (version 4.0.2)<sup>37</sup>.

## Declarations

### Acknowledgments

This work was in part supported by grants from the Japan Agency for Medical Research and Development (JP19fk0108169 to YK and JP19fk0108110/JP20he0522001 to AR).

### Author Contributions

Y. N., J. S., Y. I., T. A., and K. M. conducted proteomic analysis. K. S., R. S., M. N., and I. T. collected the samples and clinical data. M. H., Y. T., K. Y., and T. Y. conducted statistical analysis. S. K., H. K., K. M., T. K., E. H., T. O., H. H., and T. T. provided valuable suggestions for study. Data were interpreted and presented by all co-authors. S.S.J. analyzed data and wrote the manuscript. Y. K., and A. R. directed the research, analyzed the data, and wrote the manuscript.

### Competing interests

The authors declare no competing interests.

## References

1. Sohrabi, C. *et al.* Corrigendum to "World Health Organization declares Global Emergency: A review of the 2019 Novel Coronavirus (COVID-19)" [Int. J. Surg. 76(2020) 71–76]. *Int J Surg* **77**, 217, doi:10.1016/j.ijisu.2020.03.036 (2020).
2. Sanders, J. M., Monogue, M. L., Jodlowski, T. Z. & Cutrell, J. B. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*. <https://doi.org/10.1001/jama.2020.6019> (2020).
3. Chen, C. *et al.* Predicting Illness Severity and Short-Term Outcomes of COVID-19: A Retrospective Cohort Study in China. *The Innovation* **1**. 100007–100007 <https://doi.org/10.1016/j.xinn.2020.04.007> (2020).
4. Liang, W. *et al.* Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern Med*.

- <https://doi.org/10.1001/jamainternmed.2020.2033> (2020).
5. Kermali, M., Khalsa, R. K., Pillai, K., Ismail, Z. & Harky, A. The role of biomarkers in diagnosis of COVID-19 - A systematic review. *Life Sci* 254. 117788 <https://doi.org/10.1016/j.lfs.2020.117788> (2020).
  6. Whetton, A. D., Preston, G. W., Abubeker, S. & Geifman, N. Proteomics and Informatics for Understanding Phases and Identifying Biomarkers in COVID-19 Disease. *J Proteome Res*. <https://doi.org/10.1021/acs.jproteome.0c00326> (2020).
  7. Shen, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* 182, 59–72 e15, doi:10.1016/j.cell.2020.05.032 (2020).
  8. Messner, C. B. *et al.* Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst* 11, 11–24 e14, doi:10.1016/j.cels.2020.05.012 (2020).
  9. Di, B. *et al.* Identification and validation of predictive factors for progression to severe COVID-19 pneumonia by proteomics. *Signal Transduct Target Ther* 5. 217 <https://doi.org/10.1038/s41392-020-00333-1> (2020).
  10. Park, J. *et al.* In-depth blood proteome profiling analysis revealed distinct functional characteristics of plasma proteins between severe and non-severe COVID-19 patients. *Sci Rep* 10. 22418 <https://doi.org/10.1038/s41598-020-80120-8> (2020).
  11. Meftahi, G. H., Jangravi, Z., Sahraei, H. & Bahari, Z. The possible pathophysiology mechanism of cytokine storm in elderly adults with COVID-19 infection: the contribution of "inflamm-aging". *Inflamm Res* 69. 825–839 <https://doi.org/10.1007/s00011-020-01372-8> (2020).
  12. Schett, G., Sticherling, M. & Neurath, M. F. COVID-19: risk for cytokine targeting in chronic inflammatory diseases? *Nat Rev Immunol* 20. 271–272 <https://doi.org/10.1038/s41577-020-0312-7> (2020).
  13. Nishiga, M., Wang, D. W., Han, Y., Lewis, D. B. & Wu, J. C. COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. *Nat Rev Cardiol* 17. 543–558 <https://doi.org/10.1038/s41569-020-0413-9> (2020).
  14. Schijns, V. & Lavelle, E. C. Prevention and treatment of COVID-19 disease by controlled modulation of innate immunity. *Eur J Immunol* 50. 932–938 <https://doi.org/10.1002/eji.202048693> (2020).
  15. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395. 497–506 [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) (2020).
  16. Tay, M. Z., Poh, C. M., Renia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol* 20. 363–374 <https://doi.org/10.1038/s41577-020-03111-8> (2020).
  17. Sinha, P., Matthay, M. A. & Calfee, C. S. Is a "Cytokine Storm" Relevant to COVID-19? *JAMA Intern Med*. <https://doi.org/10.1001/jamainternmed.2020.3313> (2020).
  18. Kox, M., Waalders, N. J. B., Kooistra, E. J., Gerretsen, J. & Pickkers, P. Cytokine Levels in Critically Ill Patients With COVID-19 and Other Conditions. *JAMA*. <https://doi.org/10.1001/jama.2020.17052> (2020).

19. Johansen, J. S. *et al.* High serum YKL-40 level in a cohort of octogenarians is associated with increased risk of all-cause mortality. *Clin Exp Immunol* 151. 260–266 <https://doi.org/10.1111/j.1365-2249.2007.03561.x> (2008).
20. Rathcke, C. N. *et al.* Low grade inflammation as measured by levels of YKL-40: association with an increased overall and cardiovascular mortality rate in an elderly population. *Int J Cardiol* 143. 35–42 <https://doi.org/10.1016/j.ijcard.2009.01.043> (2010).
21. Furuhashi, K. *et al.* Increased expression of YKL-40, a chitinase-like protein, in serum and lung of patients with idiopathic pulmonary fibrosis. *Respir Med* 104. 1204–1210 <https://doi.org/10.1016/j.rmed.2010.02.026> (2010).
22. Lai, T. *et al.* YKL-40 expression in chronic obstructive pulmonary disease: relation to acute exacerbations and airway remodeling. *Respir Res* 17. 31 <https://doi.org/10.1186/s12931-016-0338-3> (2016).
23. Przsucha, N., Gorska, K. & Krenke, R. Chitinases and Chitinase-Like Proteins in Obstructive Lung Diseases - Current Concepts and Potential Applications. *Int J Chron Obstruct Pulmon Dis* 15. 885–899 <https://doi.org/10.2147/COPD.S236640> (2020).
24. Rajaram, S., Baylink, D. J. & Mohan, S. Insulin-like growth factor-binding proteins in serum and other biological fluids: regulation and functions. *Endocr Rev* 18. 801–831 <https://doi.org/10.1210/edrv.18.6.0321> (1997).
25. Ueki, I. *et al.* Inactivation of the acid labile subunit gene in mice results in mild retardation of postnatal growth despite profound disruptions in the circulating insulin-like growth factor system. *Proc Natl Acad Sci U S A* 97. 6868–6873 <https://doi.org/10.1073/pnas.120172697> (2000).
26. Silha, J. V. *et al.* Overexpression of the acid-labile subunit of the IGF ternary complex in transgenic mice. *Endocrinology* 142. 4305–4313 <https://doi.org/10.1210/endo.142.10.8427> (2001).
27. Wang, Z. *et al.* Insulin-Like Growth Factor-1 Signaling in Lung Development and Inflammatory Lung Diseases. *Biomed Res Int* 2018, 6057589, doi:10.1155/2018/6057589 (2018).
28. Ahasic, A. M. *et al.* IGF1 and IGFBP3 in acute respiratory distress syndrome. *Eur J Endocrinol* 166. 121–129 <https://doi.org/10.1530/EJE-11-0778> (2012).
29. Ahasic, A. M. *et al.* Predictors of Circulating Insulin-Like Growth Factor-1 and Insulin-Like Growth Factor-Binding Protein-3 in Critical Illness. *Crit Care Med* 43. 2651–2659 <https://doi.org/10.1097/CCM.0000000000001314> (2015).
30. Fan, X. *et al.* Pre-diagnostic circulating concentrations of insulin-like growth factor-1 and risk of COVID-19 mortality: results from UK Biobank. medRxiv(2020).
31. Kimura, Y. *et al.* Identification of candidate diagnostic serum biomarkers for Kawasaki disease using proteomic analysis. *Sci Rep* 7. 43732 <https://doi.org/10.1038/srep43732> (2017).
32. Ando, Y. *et al.* Nuclear pore complex protein mediated nuclear localization of dicer protein in human cells. *PLoS One* 6. e23385 <https://doi.org/10.1371/journal.pone.0023385> (2011).
33. Chenau, J., Michelland, S., Sidibe, J., Seve, M. & Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ

- labeling. *Proteome Sci* 6. 9 <https://doi.org/10.1186/1477-5956-6-9> (2008).
34. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2. 1896–1906 <https://doi.org/10.1038/nprot.2007.261> (2007).
35. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13. 731–740 <https://doi.org/10.1038/nmeth.3901> (2016).
36. von Mering, C. *et al.* STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 33. D433–437 <https://doi.org/10.1093/nar/gki005> (2005).
37. Kanda, Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplant* 48. 452–458 <https://doi.org/10.1038/bmt.2012.244> (2013).

## Tables

Due to technical limitations, table 1 is only available as a download in the Supplemental Files section.

## Figures

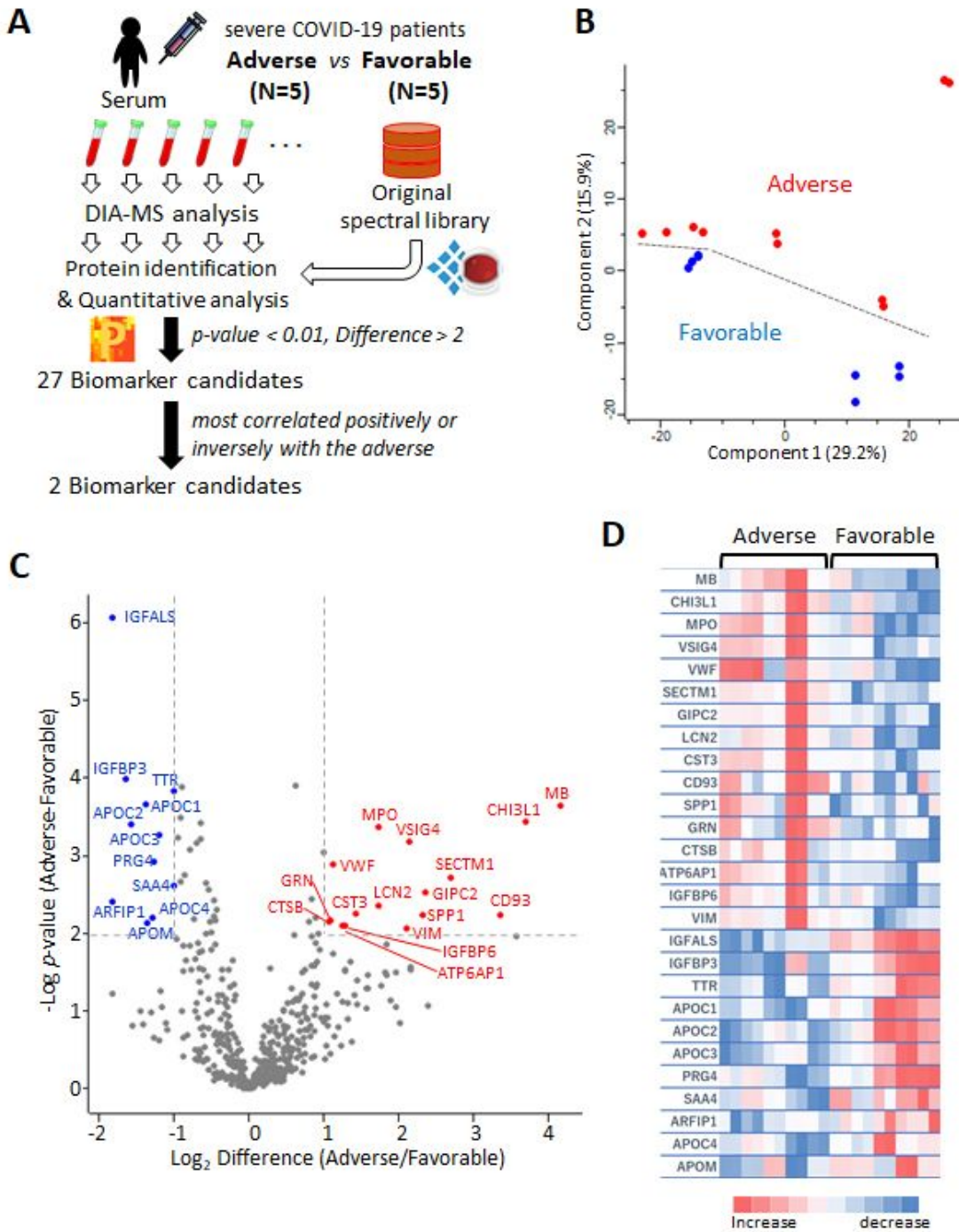


Figure 1

Figure 1

Serum protein profile analysis of critical COVID-19 patients with adverse and favorable prognosis by DIA-MS analysis A, Study design for identification of potential biomarker candidates. Serum protein profiling obtained from the discovery study comprising five adverse prognosis and favorable prognosis by DIA-MS analysis. B, PCA score plot for proteins in two groups with Perseus are shown here. Critical patients with adverse and favorable prognosis are represented by red and blue, respectively. C, Volcano plot

representing the difference in serum expression levels of 495 proteins between the two groups. Red or blue dots indicate proteins with increased or decreased serum expression levels in patients with adverse prognosis compared to those with favorable prognosis, respectively ( $p$ -value < 0.01, difference [adverse/favorable] > 2) (Listed in Table 1). D, Heatmap visualization of the 27 differentially expressed proteins between severe COVID-19 patients with adverse and favorable prognosis. Increased and decreased are in respect to adverse vs. favorable. Gene names are shown on the left.

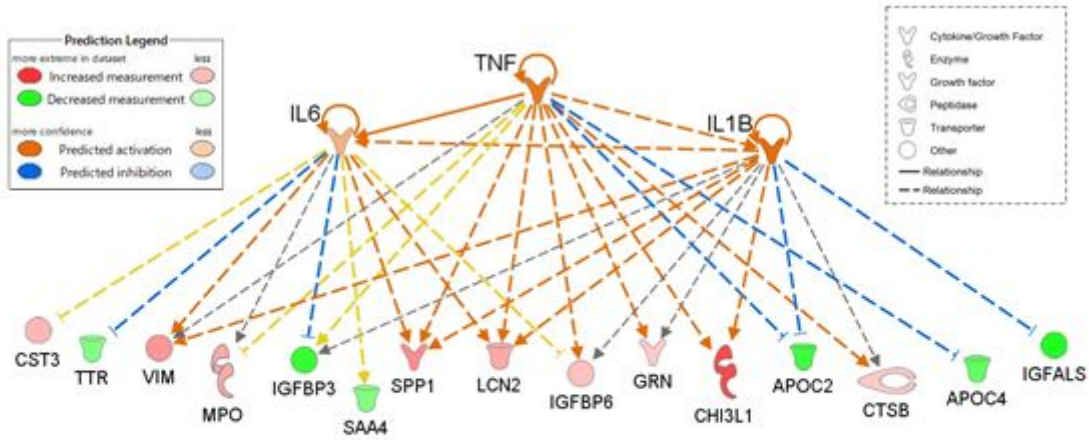


Figure 2

Figure 2

Network of proteins by the biological upstream analysis within the framework of the IPA. Solid and dotted lines indicate direct and indirect interaction, respectively. Orange nodes represent molecules undetected in our study.

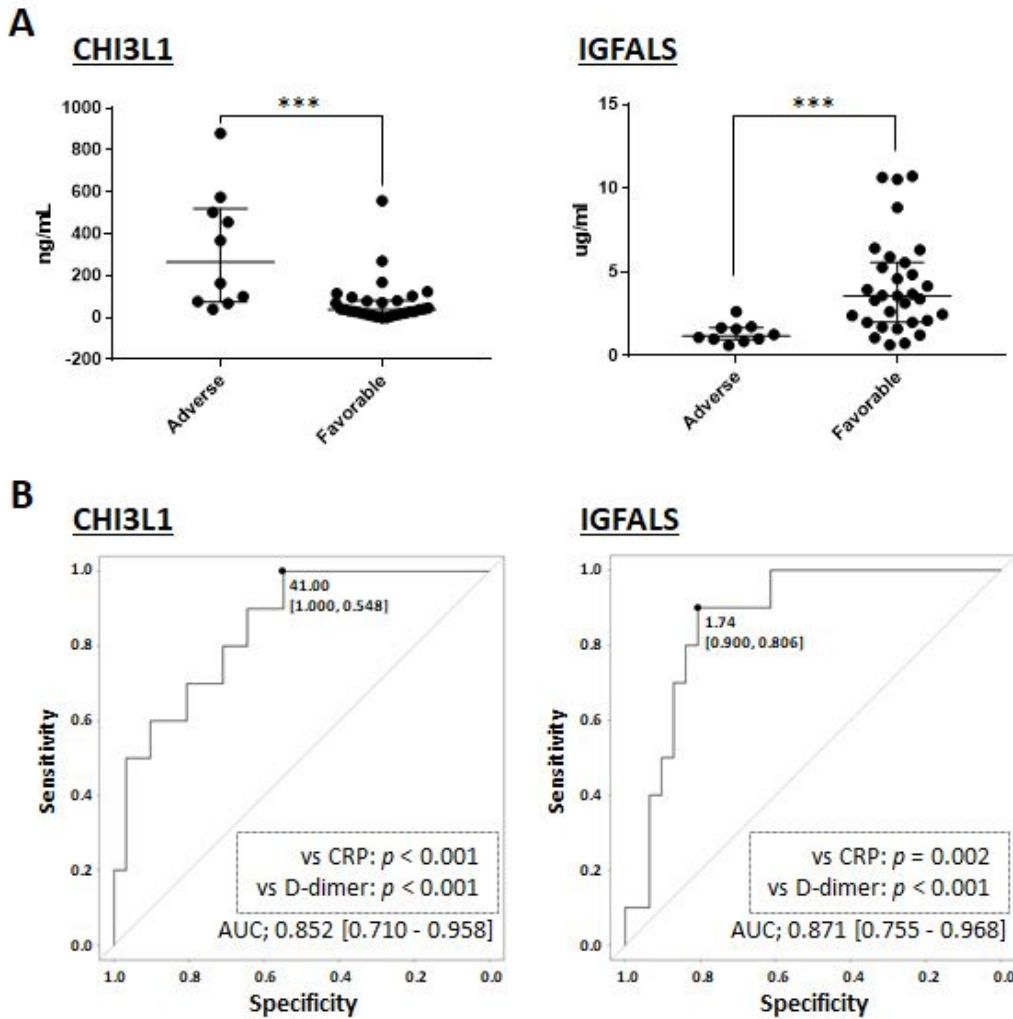


Figure 3

Figure 3

Serum levels and ROC curves of CHI3L1 and IGFALS with ELISA assay A, In 10 COVID-19 patients with adverse prognosis, the median serum levels [interquartile range, IQR] of CHI3L1 and IGFALS were 266.52 [83.43, 491.76] ng/mL and 1.19 [1.01, 1.67]  $\mu$ g/mL, respectively. In 31 COVID-19 patients with favorable prognosis, the median serum levels [IQR] of CHI3L1 and IGFALS were 38.07 [18.82, 80.19] ng/mL and 3.59 [2.06, 5.42]  $\mu$ g/mL, respectively. Differences of continuous values between two groups were



compared by the Mann-Whitney U-test: \*\*\*;  $p < 0.001$ . B, ROC curve for prediction of clinical prognosis of severe COVID-19 patients. The optimal cut-off values [sensitivity and specificity] for making a clinical decision using CHI3L1, or IGFALS were defined as 41.00 ng/mL [1.000 and 0.548] and 1.74  $\mu\text{g/mL}$  [0.900 and 0.806] respectively. The p-value is generated from bootstrap test for two ROC curves. AUC indicates the area under the ROC curve [95% CI].

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.xlsx](#)
- [SupplementalFigure.pptx](#)
- [SupplementalTable.xlsx](#)