

Robust prognostic signatures derived from proteomic panel profiling of plasma from patients with severe COVID-19 complications

Maryam Al-Nesf

Hamad Medical Corporation <https://orcid.org/0000-0001-9354-0214>

Houari Abdesselem

Qatar Biomedical Research Institute

Ilham Bensmail

Qatar Biomedical Research Institute

Shahd Ibrahim

Hamad Medical Corporation <https://orcid.org/0000-0003-0815-4840>

Walaa Saeed

Hamad Medical Corporation <https://orcid.org/0000-0001-7342-7951>

Sara Mohammed

Hamad Medical Corporation

Almurtada Razok

Hamad Medical Corporation

Hashim Alhussain

Qatar University

Reham Aly

Qatar University

Muna Al Maslamani

Hamad Medical Corporation

Khalid Ouararhni

Qatar Biomedical Research Institute

Mohamad Khatib

Hamad Medical Corporation

Ali Ait Hssain

Hamad Medical Corporation

Ali Omrani

Hamad Medical Corporation <https://orcid.org/0000-0001-5309-6358>

Saad Al-Kaabi

Hamad Medical Corporation

Abdullatif Al Khal

Hamad Medical Corporation

Asma Althani

College of Arts and Sciences/ Qatar University

Waseem Samsam

Anti-Doping Laboratory Qatar

Abdulaziz Farooq

Aspetar Hospital

Jassim Al-Suwaidi

Hamad Medical Corporation

Mohammed Al-Maadheedh

Anti-Doping Laboratory Qatar

Heba Al-Siddiqi

Qatar Biomedical Research Institute

Alexandra Butler

Qatar Biomedical Research Institute

Julie Decock

Qatar Biomedical Research Institute

Vidya Mohamed-Ali

Anti-Doping Laboratory Qatar

Fares Al-Ejeh (✉ falejeh@hbku.edu.qa)

Qatar Biomedical Research Institute <https://orcid.org/0000-0002-1553-0077>

Article

Keywords: COVID-19, SARS-CoV-2, plasma

DOI: <https://doi.org/10.21203/rs.3.rs-126482/v1>

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

[Read Full License](#)

Abstract

COVID-19 complications present a huge burden on healthcare systems and warrant a predictive risk model for disease severity of SARS-CoV-2 infection to enable early intervention, prospective decision-making and triaging of patients. We profiled plasma proteins from COVID-19 patients (severe n=50, and mild n=50) and controls (n=50) using function- and pathway-based panels developed with the highly specific proximity extension assays. Several biological pathways were specific for patients with severe complications. Based on these dysregulated profiles, we propose candidate FDA-approved drugs that target multiple upregulated proteins to treat severe complications. In addition, the set of differentially expressed plasma proteins in severe disease contained a robust 46-protein signature, the COVID-19 molecular severity score, which predicts the risk of severe complications. We cross-validated this molecular severity score in an independent cohort and found it useful within three days after hospital admission to predict COVID-19 severity and outcomes. Associated with the molecular severity score, we identified a set of clinical parameters available at admission, that act as a clinical risk score for complications. The molecular and clinical risk scores described in our study may be prognostic tools for severe COVID-19 disease and help alleviate the pressure on healthcare systems during infection peaks.

Introduction

The rapid and wide-spread dissemination of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has pressured, and tested, healthcare systems globally. To date, there have been over 60 million individuals infected worldwide, leading to over 1.5 million deaths due to severe complications from the Coronavirus disease 2019 (COVID-19).

To date, the most comprehensive report was released by the International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC) on 3 September 2020 and included data for 96,074 individuals from 562 sites across 42 countries. The ISARIC reported 81,705 laboratory-confirmed SARS-CoV-2 infections (95.8%) of which 46,516 (56.9%) were observed in males and 35,033 (42.9%) in females, with a median age of 72 years. The most common comorbidities seen at admission included cardiovascular disease (CVD, 31.4%), hypertension (HTN, 48%), diabetes (17.2%), chronic kidney disease (CKD, 16.9%), and chronic pulmonary disease (16.8%). History of fever, shortness of breath, cough, fatigue, and confusion were the most common symptoms on admission. Oxygen saturation (SpO₂%) less than 94% was present in 31.3% of the patients. Admission to intensive care or high dependency units (ICU/HDU) at some point of illness, which could be defined as severe COVID-19 complications, was reported for 13,977 (17%) patients and of those 4,770 (34.1%) died while 6,348 (45.4%) were discharged by the time of the ISARIC report (1).

While several studies reported on symptoms and comorbidities that associate with severe COVID-19 complications, early prognostic tools to stratify the risk of developing complications are still lacking. In this study, we hypothesized that changes in plasma proteins offer prognostic molecular profiles that can also identify the most informative clinical features presented at admission to predict the risk of

developing complications. To address this, we used proteomic panel-profiling of plasma from patients with severe COVID-19 complications versus mild-moderate symptoms and control subjects to characterize biological processes and pathways associated with disease severity. Then we evaluated the clinical findings that associated with the molecular changes.

Results

Study cohort characteristics

Characteristics of the study groups, patients and controls, are summarized in **Table 1**. Most infected patients were males ($n = 91$, 91%) and severe COVID-19 disease, defined by admission to ICU, was associated with age ($p < 0.001$), particularly in patients over 40 years of age in our cohort. The distribution of ethnicity in the severe and mild-moderate groups was not significantly different, but the control group had a higher percentage of South Asians ($p = 0.035$). Patients with severe disease had a significantly higher BMI and were either overweight ($n = 25$, 50%) or obese ($n = 18$, 36%), and had a significantly higher heart rate and lower oxygen saturation (SpO_2). Moreover, diabetes and hypertension were significantly associated with severe complications in the lungs and kidneys, compared to mild-moderate disease (**Table S1**).

High differential protein expression in plasma from patients with severe complications

Plasma samples from 100 SARS-CoV-2 infected patients, 50 with severe and 50 with mild-moderate disease, and 50 control subjects were analyzed using ten different 92-protein Olink panels (**Table S2**). For one patient, P064, Olink assays failed QC for seven out of the ten panels thus this case was excluded. The number of differentially expressed proteins (DEPs) from single panel analyses for samples which passed Olink's QC (**Figure S1**) is summarized in **Figure 1A**. Severe disease versus control showed many DEPs, more than 46 out of 92 (50%) proteins per panel, whereas the number of DEPs in each panel was less when comparing severe vs. mild and mild vs. control. The cardiovascular panels had the most DEPs in patients with severe complications compared to patients with mild-moderate disease and controls. Receiver operating characteristic (ROC) curve analyses using all the DEPs, calculated as a single score for each panel, found that all panels had a significant predictive value, high areas under the curve (AUC), to identify severe cases ($p < 0.01$) (**Figure 1A**).

For a comprehensive molecular view, we carried out the analysis after combining the data from the ten Olink panels in a single dataset (**Table S2**). Unsupervised hierarchical clustering, before any filtering, revealed that the ten panels could differentiate severe disease from mild-moderate disease and controls (**Figure 1B**). A larger number of DEPs was identified when severe disease was compared to mild-moderate disease or controls than when comparing mild-moderate disease to controls (**Figure 1C, Table S3**). The DEPs found in the combined analysis included all but 18 out of the 567 (3%) DEPs identified by the single

panel analysis which did not meet the statistical significance in the combined analysis (adjusted p-value <0.05 with FDR <0.1). It should be noted that the DEPs in severe disease were mainly upregulated, whereas DEPs in mild disease versus controls had equal distribution of up- and down-regulation (**Figure 1C**). This significant trend (Chi-square p-value <0.0001) indicated a disruption of biological responses measured by changes in systemic proteins in patients with severe COVID-19 complications.

Functional analysis of the deregulated proteins in plasma of infected patients

The DEPs in severe disease versus mild disease and controls, and mild versus controls, were subjected to KEGG pathways enrichment analysis. The statistical significance of enriched pathways should be treated cautiously since our proteomic assays were based on enriched panels consisting of 894 unique proteins; however, relative enrichment is warranted. Cytokine-cytokine receptor interaction were enriched gradually from control subjects to patients with mild to severe diseases (**Figure S2A-C**). Such gradual enrichments were observed for several pathways that mainly relate to immune, inflammation and infection-related pathways, and the associated cell signaling pathways. The three comparisons, severe versus mild-moderate or controls and mild-moderate versus control, had 72 upregulated and 22 downregulated overlapping proteins (**Figure S2D**). Additionally, 16 and 56 upregulated and downregulated proteins, respectively, were common in severe and mild-moderate disease compared to controls. Protein-protein interaction (PPI) analysis of these common 166 DEPs revealed a highly connected network with an average of 11 connections per protein (data not shown).

Next, we dissected the differences between patients with severe versus mild-moderate disease by excluding the significantly different DEPs between mild-moderate cases versus controls; changes specific to severe COVID-19 complications. As shown in **Figure 2A**, most of the 106 DEPs in severe versus mild disease were also significantly deregulated compared to controls and were highly connected based on PPI analysis (average 13 connections per protein). The PPI analysis revealed a clear pathway starting from cytokine-cytokine interaction and chemokine signaling, particularly IL-17 signaling, to downstream signaling pathways including TNF, NF κ B, Jak-STAT, PI3K, Ras and Rap1 signaling (**Figure 2B**). Three additional pathways were also identified in this analysis including complement and coagulation, cell adhesion, and apoptosis and lysosomal degradation.

Potential drugs to target deregulated proteins in COVID-19 patients with severe complication

In addition to targeting the enriched pathways (**Figures 2 and S2**) such as TNF α , coagulation or Jak-STAT, an analysis of protein-drug interaction (PDI) was carried out based on the upregulated proteins in patients with severe versus mild disease. A library of FDA-approved drugs was screened for interactions with the significantly upregulated proteins (> 1.5-fold) and identified 263 drugs which we grouped according to

drug classes (**Table S4**) and summarized in overlapping PPI and PDI networks (**Figure S3A**). Briefly, the FDA-approved drugs included anti-viral drugs (e.g. ribavirin and ritonavir) and anti-coagulant and thrombolytic drugs (e.g. dalteparin, tenecteplase and vorapaxar). The list also included corticosteroids and glucocorticoids (e.g. dexamethasone, methylprednisolone and prednisone), nonsteroidal anti-inflammatory drugs (NSAID, e.g. aspirin and indomethacin), nonsteroidal antiandrogen (NSAA, e.g. bicalutamide and flutamide), and immunosuppressive and immunomodulatory drugs (e.g. anakinra methotrexate filgrastim, tocilizumab, atezolizumab, and nivolumab).

Several drugs interact with proteins that were upregulated by 2-fold or more in severe versus mild-moderate disease (**Figure 3** and **Figure S3B**). One PDI network showed high connectivity with few drugs interacting with 2 or 3 protein targets (**Figure 3A**). We selected those PDIs where the drugs interacted with more than one protein and considered the direction of protein-protein interactions within the PDI network (**Figure 3B**). Hepatocyte Growth Factor (HGF) and Myeloperoxidase (MPO) had the top out-interactions (effector) with 12 and 11 outward interactions, respectively, followed by CXCL10 which had seven out-interactions (**Figure 3A&B**). Hence, drugs targeting HGF, MPO and CXCL10 are expected to influence most of the network's interactions. Flutamide, a nonsteroidal antiandrogen, can target ACE2 and MPO which were upregulated more than two-fold, and IL2RA which showed 1.7-fold upregulation in patients with severe disease. The immunomodulatory drug thalidomide is another example that targets four upregulated proteins: IL6R, Von Willebrand factor (VWF) and HGF with more than two-fold upregulation and F2R with 1.5-fold upregulation in severe disease. Other examples are outlined in Discussion.

The molecular severity score: a 46-protein signature for COVID-19 severity

To develop a blood-based protein signatures we used the MUVr tool (2) for variable selection and validation in multivariate modelling to identify the most stable DEPs that can differentiate all groups (patients and control subjects), severe versus mild disease, or severe or mild disease versus controls. Four predictive models, MUVr modules, were identified through selecting the minimum number of DEPs with the least frequency of misclassifications (**Figure 4A**). Module i had 10 DEPs which were common in three out of the four analyses whereas module ii had 36 DEPs that were shared in the analysis of all patients and the comparison of severe versus mild disease. Modules iii and iv consisted of DEPs unique for all patients and patients with severe versus mild disease, respectively. ROC curves found that each of the four modules had a statistically significant predictive potential ($p < 0.0001$) with high AUC (**Figure 4B**). All modules except module iii could differentiate patients with severe disease versus controls with 100% specificity and sensitivity. For patients with severe disease versus patients with mild disease and controls, modules ii and iv had AUC of 1 (100% specificity and sensitivity). Modules ii and i were the best predictors of severity (**Figure 4B**), thus these two modules, consisting of 46 DEPs, were combined to calculate the "COVID-19 molecular severity score". The molecular severity score had 100% sensitivity and specificity for all comparisons except for patients with mild disease versus controls; however, the significant predictive capacity endured (**Figure 4C**).

Interestingly, although variable selection in MUVR is blinded to any biological information, the selected proteins had high interactions forming a solitary connected network from 37 (80%) out of the 46 DEPs (**Figure 4D**). Gene ontology biological process (GO-BP) enrichment analysis of this 46-protein signature identified the same enriched processes described earlier (**Figure 2 and S2**), indicating that this minimum set of DEPs recapitulated the larger differential profiles. The complete GO-BP enrichments can be classified into five broad processes: viral entry, signaling, immune response, metabolism, and development (**Table S5 and Figure S5**).

The COVID-19 molecular severity score validates in an independent cohort

To validate the COVID-19 molecular severity score (46-protein signature), we calculated this score for patients in the independent Massachusetts General Hospital (MGH) cohort (3). As shown in **Figure 5A**, the molecular severity score was significantly higher in patients with severe disease (Acuity 1 - death and Acuity 2 – intubated, ventilated but survived 28 days) compared to non-severe disease (Acuity 3-5). This was particularly clear when evaluating the maximum acuity over the 28-day study period (Acuity Max) as illustrated in the time curve shown in **Figure 5B**. The molecular severity scores from plasma samples collected on days 3 and 7 were highly predictive of COVID-19 severity (death or intubation/ventilation but survived, or both, **Figure 5C**). Pairwise comparison of showed that plasma samples from days 3 and 7 were more useful than day 0 (**Figure 5D**).

It should be noted that the MGH study included symptomatic patients who were SARS-CoV-2 negative (n = 78). The molecular severity score on day 0 was significantly higher in SARS-CoV-2 positive patients who died or were intubated for mechanical ventilation (Acuity groups 1 and 2) compared to SARS-CoV-2 patients (**Figure S5A**). However, the molecular severity scores were not different between non-severe SARS-CoV-2 positive and negative patients (Acuity groups 3-5). Moreover, the molecular severity score on day 0 showed significant prediction of disease severity in symptomatic patients irrespective of SARS-CoV-2 infection status, albeit the prediction is more accurate for SARS-CoV-2 positive patients (**Figure S5B**). These results suggest that the molecular severity score reflects a general biological response in symptomatic respiratory infection or disease but was more prominent in SARS-CoV-2 positive patients.

A molecularly trained clinical score to predict COVID-19 severity

We hypothesized that the molecular severity score could be used to identify informative clinical parameters to triage SARS-CoV-2 infected patients into high or low risk for developing severe complications. A comparative analysis between the molecular severity scores and clinical parameter found that 13 out of the 24 parameters available in our cohort showed significant associations with the molecular severity score (**Figure 6S**). MUVR determined that 8 of these are the most informative (**Figure**

6A). These eight parameters were combined to calculate the “Clinical Risk Score”; a molecularly trained score where each clinical measure was weighted according to its molecular severity score. The Clinical Risk Score outperformed each of the single parameters ($p < 0.0001$ for all pairwise comparisons of ROC AUCs) and showed more than 90% sensitivity and specificity to differentiate between patients with severe and mild disease (**Figure 6B**). Of the eight parameters used in the Clinical Risk Score, three were also available in the MGH cohort; neutrophil and lymphocyte counts and CRP levels. Each of these three parameters correlated with the molecular severity score in the MGH cohort, and combining these parameters with the same cutoffs defined in our cohort (**Figure 6A**) showed strong predictive value of COVID-19 severity in the MGH cohort (AUC 0.816, sensitivity 84.5% and specificity 68.3%). In conclusion, both the 46-protein severity score and the trained clinical score may be used during admission to identify SARS-CoV-2 infected patients who are at risk for developing severe complications. Moreover, as shown in **Figure 6C-D**, several highly up- and down-regulated proteins may be used for molecular monitoring of disease progression or response to treatments and could serve as candidate therapeutic targets for intervention.

Discussion

Demographics of the current cohort of SARS-CoV-2 infected patients had similar characteristics as did the nationwide cohort study of the first consecutive 5,000 patients with COVID-19 in Qatar (4). While our study selected patients between 18 and 65 years of age, both cohorts consisted largely of males with a younger median age due to the relatively younger population in Qatar. Risk factors of ICU admission in our study and the national cohort study (4) included older age, male sex, higher BMI, and preexisting diabetes and hypertension. Other comorbidities such as chronic artery disease, liver disease or kidney disease were identified in both studies but did not reach significance in our smaller cohort. Of the first 5,000 consecutive cases in Qatar, 1424 patients (28.5%) required hospitalization, out of which 108 (7.6%) were admitted to ICU, and only 14 patients (0.28%) had died by 60 days after infection. In a relatively younger national cohort in Qatar, with a low comorbidity burden, COVID-19 was associated with low mortality (4), which was also reflected in our smaller cohort.

The Qatari population, with respect to COVID-19, is unique in the demographic characteristics (e.g. predominantly males with younger age) when compared to other populations such as that described in the ISARIC (1). However, our study of plasma profiling of SARS-CoV-2 infected patients and control subjects using the Olink Proteomics panels may be generalizable as the biological processes and pathways enriched in patients with severe complications in our subpopulation have also been reported in other studies. Additionally, the COVID-19 molecular severity score reported here was cross-validated in an independent, ethnically different, larger cohort from the Massachusetts General Hospital (MGH, USA). Interestingly, the analysis in the MGH cohort which included symptomatic SARS-CoV-2 negative patients showed that the molecular severity score could reflect a general biological response in symptomatic respiratory infection or disease irrespective of SARS-CoV-2 status although the score was higher and more predictive in SARS-CoV-2 positive patients.

Medical history of macular degeneration and of coagulation disorders (thrombocytopenia, thrombosis, and hemorrhage) were considered risk factors for higher morbidity and mortality in a recent study on 11,116 patients infected with SARS-CoV-2 (5). Moreover, RNA-Seq profiles from nasopharyngeal swabs in this study found several enriched immune-modulatory functions in SARS-CoV-2 infected patients versus controls, such as inflammatory response, interferon alpha response, and IL6-JAK-STAT3 signaling, which were also identified in our study. Activation of the complement and coagulation cascades was also among the most enriched gene sets (5), an observation corroborated by our plasma protein profiling results.

A multiplexed biomarker profiling of plasma from 49 SARS-CoV-2 infected patients (40 in ICU and 9 in non-ICU units) and 13 non-COVID-19, non-hospitalized controls identified multiple proteins in association with ICU admission and mortality, including HGF, RETN, LCN2, G-CSF, IL-6, IL-8, IL-6, IL-10, IL1RA and TNF- α (6), which were confirmed in our study. Importantly, the study also reported a unique neutrophil activation signature composed of neutrophil activators (G-CSF, IL-8) and effectors (RETN, LCN2 and HGF), with a strong predictive value to identify critically ill patients whereby the effector proteins strongly correlated with absolute neutrophil count (6). Our study not only identified those components of the neutrophil activation signature, but also found that the COVID-19 molecular severity score, a more comprehensive signature, also correlated with absolute neutrophil counts. Moreover, the neutrophil count was selected in the COVID-19 clinical risk score developed in our study. Another study deployed Olink Proteomics panels to measure 1,161 plasma proteins from 20 patients, 10 SARS-CoV-2 positive and 10 SARS-CoV-2 negative patients, admitted to ICU and 10 healthy controls (7). This study had a small sample size, therefore could not determine changes contributing to ICU admission and only reported mortality. Interestingly, it uncovered similar proteins and pathways as those identified in our study in association with COVID-19 severity, such as interleukins, CXCLs/chemokines, membrane receptors linked to lymphocyte-associated microparticles, cytoplasmic/cytoskeletal proteins, and nuclear proteins or transcription factors (7). Among their reported 20 top proteins differentiating patients with COVID-19 disease from healthy controls, 13 (65%) were also confirmed in our study, with two of them being components of our COVID-19 molecular severity score, namely IL6 and IL18R1. Of their reported 20 top proteins which differentiated ICU-admitted patients with COVID-19 versus non-COVID-19 disease, 12 (60%) were also found in our study with two being components of the molecular severity score, KRT19 and CCL7.

Besides Olink technology, mass spectroscopy was used in two studies to identify deregulated proteins in SARS-CoV-2 infected patients. The first study used liquid chromatography-mass spectrometry (LC-MS) to profile 31 patients with SARS-CoV-2 infection, where the disease severity was graded according to the WHO outcome scale. The study identified 27 potential biomarkers that were differentially expressed (8). Although none of these biomarkers were identified in our study, the biological functions reported in their study were also captured in our analysis, including complement factors and the coagulation system, inflammation modulators, and pro-inflammatory factors upstream and downstream of interleukin 6. The second study used mass spectroscopy for proteomic and metabolomic analysis of sera from 46 COVID-19 and 53 control individuals and identified 93 proteins which were differentially expressed in sera from

severe COVID-19 patients (9). Of these 93 differentially expressed serum proteins, 17 were included in our panel profiling of plasma with 11 (65%) of these proteins also being significantly deregulated in our cohort. More specifically, 8 of the 11 proteins were upregulated (VWF, PVR, GRN, NID1, VCAM1, SAA4, CD59, CDH1), whereas the remaining three proteins were downregulated (FETUB, APOM, and IGFBP3) in the severe cases in our study. Using mass spectroscopy, the authors developed a model of 22 proteins and seven metabolites (29 sera factors) to stratify patients according to severity (9). None of the protein biomarkers reported in this study were identified in our study; however, there is a strong overlap in biological functions, including the release of IL-6 and TNF- α , inflammatory responses, activation of the complement system and protein phosphorylation.

Interestingly, both MS-based studies agreed on ten protein biomarkers in their classifiers, 10/27 (37%) for the first study which included serum and plasma (8), and 10/22 (45%) in the second study which only used serum (9). Although our study agrees on the biological functions identified in the two MS-based studies, the lack of agreement with the named biomarkers may be due to the use of plasma in our study compared to serum in the MS studies. We cannot exclude that the MS-based studies are more comprehensive and less biased than the panel profiling used in our study. However, it should be noted that there was a small overlap between all the proteins detected by mass spectroscopy in sera (prior to statistical analysis) from the Shen et al. study (9) and the proteins profiled in plasma in our study; 134 common proteins out of the 791 (17%) proteins detected by mass spectroscopy and the 894 (15%) proteins profiled in our study.

Altogether, our study identified several biological pathways described in previous proteomic studies of sera or plasma of patients with severe COVID-19 complications. In addition to their potential biomarker value, the protein profiles can also be used to predict potential drugs for intervention. Our drug-protein interaction analyses shortlisted HGF, MPO and CXCL10 as targets that could influence most of the interactions between the plasma proteins upregulated in severe COVID-19 cases. Notable examples of possible drugs include flutamide which can target MPO, ACE2 and IL2RA and has been proposed as a possible drug for COVID-19 treatment based on ACE2 interaction network analysis (10). Methylene Blue can modulate MPO, VWF and CPA1, which were upregulated in our severe cases and had been tested in combination with other drugs in a clinical trial with critically ill COVID-19 patients in Iran (11), whilst a wider clinical trial (NCT04370288) has been designed. Thalidomide is another example that targets one of the shortlisted proteins (HGF) in addition three other upregulated proteins in severe cases (IL6R, VWF, and F2R). Its use for COVID-19 was reported for a single case in China (12) and led to recovery, and two clinical trials (NCT04273581 and NCT04273529) have been registered. However, thalidomide's side effects and its previous dark past has been raised as serious concerns for its use to treat COVID-19 patients, and may have to be strictly limited to use in men and post-menopausal women (13).

Furthermore, methotrexate inhibits HGF and two other upregulated proteins in severe cases in our study, S100A12 and SULT2A1. This drug has been reported to inhibit SARS-CoV-2 virus replication *in vitro* via purine biosynthesis, thereby potently inhibiting viral RNA replication, viral protein synthesis, and virus release. As such, methotrexate was proposed as an effective measure to prevent possible COVID-19

complications (14). The use of methotrexate to treat COVID-19 patients or prevent complications has not been tested; however, a large comparative cohort study suggested that patients with recent TNF inhibitors and/or methotrexate exposure do not have increased COVID-19 related hospitalization or mortality (15).

In addition to the potential targeting of HGF, MPO and CXCL10 as highly interconnected proteins, our analysis identified ribavirin as a treatment option based on the upregulation of VWF and CST3 (Cystatin C) in patients with severe COVID-19 complications. Ribavirin, an oral nucleoside analogue, has been tested in combination with injectable interferon beta-1b and the oral protease inhibitor (lopinavir-ritonavir) in a randomized phase 2 trial to treat COVID-19 patients. Compared to lopinavir-ritonavir alone, the triple combination was safe and effectively shortened the duration of virus shedding, decreased cytokine responses, alleviated symptoms, and facilitated the discharge of patients with mild to moderate COVID-19 disease (16). A follow-up trial has been registered (NCT04494399) to test the combination of ribavirin with interferon beta-1b, without lopinavir-ritonavir, to treat patients with COVID-19.

In conclusion, our study identified deregulated proteins in the plasma of patients with severe COVID-19 complications that may inform therapeutic interventions. The 46-protein signature identified in our study was developed as the COVID-19 molecular severity score and used to stratify patients according to COVID-19 severity in an independent cohort. The COVID-19 molecular severity score could predict outcomes up to 28 days post-admission and from as early as three days of admission. We used the molecular severity score to select clinical parameters available at the time of admission and generated a scoring system to develop the molecularly trained clinical risk score. The molecular severity and the clinical risk scores developed here have the potential to stratify SARS-CoV-2 infected patients at early stages according to their risk of developing complications to prospectively inform healthcare management and clinical decision-making to prevent complications and mortality.

Methods

Patients recruitment

A cohort of 100 patients (mild-moderate and severe) affected by COVID-19 disease and admitted to Hamad Medical Corporation (HMC) hospitals; tertiary level hospitals in Doha, Qatar, were recruited. Infection was confirmed by positive RT-PCR assays for SARS-CoV-2 from sputum and throat swab with CT values around 30. Patients with severe COVID-19 were defined as those requiring ICU admissions due to COVID-19 disease or disease complications, while patients with mild-moderate COVID-19 were admitted to community hospitals but did not requiring ICU care. Fifty control subjects were recruited at the Clinical Research Center of the Anti-Doping Laboratory Qatar from volunteers identified by Qatar Red Crescent Society, according to the criteria of being healthy, without prior history of confirmed COVID-19 infection diagnosis, normal oxygen saturation, and vital signs. Control subjects were age, sex and ethnicity matched to the patients. Individuals with poor cognitive ability, or any past or present medical disease or were not able to consent were excluded.

Samples collection and processing

Peripheral blood was collected within five to seven days of admission into commercially available EDTA-treated tubes, and plasma and peripheral blood mononuclear cell (PBMC) fractions were separated using Ficoll. PBMCs were saved for use in other studies. Plasma was stored at -80°C until further analysis.

Olink proteomic assays

Plasma samples were profiled in-house using the proximity extension assays (PEA), 96-plex immunoassay developed by Olink Proteomics (Uppsala, Sweden) (17) following the standard protocol at Qatar Biomedical Research Institute's (QBRI) Olink certified proteomics core facility. Quality control and data normalization according to internal and external controls were carried out using the Normalized Protein eXpression (NPX) software and every run was checked and validated by the Olink support team in Uppsala. Ten different panels, each focused on a specific disease or biological process, were used in our study; panel names are stated in the results.

Bioinformatics

For the analysis of Olink assays, the protein expression values, as log₂ of Normalized Protein eXpression (NPX), were used. Two approaches were used in the analysis; single-panel and combined-panels analyses before confirming the overlap between the two approaches. Olink data that did not pass quality control was excluded from the analyses. R packages for hierarchical clustering (heatmap.2), principal component analysis (PCA, prcomp), differential expression analysis (Linear Models for Microarray Data (limma)), volcano plots, gene-ontology biological process (GO-PB) and KEGG pathways enrichment analyses were used through the standalone version of iDEP.92 (18) installed in RStudio (version 1.2.5).

For variable selection and validation, the algorithm for multivariate modeling with minimally biased variable selection in R (MUVR) was used in RStudio as previously described (2). MUVR is a statistical validation framework, incorporating a recursive variable selection procedure within a repeated double cross-validation (rdCV) scheme. Differentially expressed proteins selected by MUVR were used to develop protein signatures represented as meta-protein scores calculated as the ratio of average expression of NPX values of upregulated proteins to the average expression of NPX values of downregulated proteins. Upregulated and downregulated proteins were defined according to the score. For example, if the score was from the comparison of severe versus mild COVID-19 patients, we used the upregulated or downregulated proteins in the severe versus mild patients. Scores were evaluated using receiver operating characteristic (ROC) curve analyses to determine the area under the ROC curve (AUC), sensitivity, specificity, and significance ($P < 0.05$) using MedCalc® (version 12.7, MedCalc Software Ltd., Belgium).

Protein-protein interaction (PPI) was analyzed and visualized using the STRING database (STRING-db version: 11.0) (19) accessed through Cytoscape (version: 3.7.2) (20). Protein-drug interaction (PDI) was

analyzed using the Drug-Gene Interaction database (DGldb, v3.0.2) (21), only using FDA-approved drugs, and interaction networks were visualized in Cytoscape.

Validation of the COVID-19 molecular severity score in the MGH cohort

To validate the COVID-19 molecular severity score (46-protein signature) developed here, we used the Massachusetts General Hospital (MGH) cohort. The MGH cohort enrolled 384 acutely-ill patients, 18 years or older patients, with a clinical concern for COVID-19 upon arrival in the emergency department as described previously (3). SARS-CoV-2 positivity was reported for 306 patients (80%) while the remaining 78 patients were SARS-CoV-2 negative. The COVID-19 molecular severity score was calculated as described above (meta-protein score) for each patient. The performance of the COVID-19 molecular severity scores in the MGH cohort was evaluated with ROC curve analysis using MedCalc® (version 12.7, MedCalc Software Ltd., Belgium).

Statistics

Patient clinical data analysis was performed using Statistical Package for Social Sciences (SPSS v26, Chicago IL, USA). Groups were compared using the chi-square test, and Fisher's exact test (two-tailed) replaced the chi-square in the case of a small sample size where the expected frequency is less than 5 in any group. The results were presented as mean \pm SD for normally distributed data or median (IQR) for skewed results and/or number and percentage of participants as appropriate. The level of statistical significance was set at $p < 0.05$. GraphPad Prism (version 8.4.3, GraphPad Software LLC, CA, USA) was used to compare protein signature scores across clinical subgroups using unpaired, two-tailed t-tests or one-way ANOVA with Dunnett's multiple testing correction.

Study approval

The study received IRB approval from the Hamad Medical Corporation (HMC, Doha - Qatar) and was supported by a grant from HMC-Medical Research Council (MRC); approval and fund number MRC-05-003. Written informed consent was obtained from all the participants in the study. The conduct of this study was in accordance with ICH GCP and the Declaration of Helsinki.

Declarations

Author contributions

Conception: M.AY.A, H.B.A., J.A., M.Al-Maadheed, H.H.A., A.E.B., J.V.D., V.M-A, F.A. Designing the study: M.AY.A, H.B.A., M.Al Maslamani, M.Y.K, A. Ait Hssain, A.S.O., S.A., A. Al Khal, A.A.A., J.A., M.Al-Maadheed,

H.H.A., A.E.B., J.V.D., V.M-A, F.A. Conducting experiments and acquiring data: M.A.A, H.B.A., I.B., S.I., W.AH.S., S.SI.M., A.R., H.A., R.MA.A., K.O., W.S. Analyzing the data: M.AY.A, H.B.A, A.F., F.A. Writing the manuscript: M.AY.A, H.B.A., H.H.A., A.E.B. J.V.D., V.M-A., F.A., Revision and editing the manuscript: All authors.

Acknowledgment

The authors would like to thank all the patients, volunteers and the healthcare co-workers from Allergy and Immunology Section-HMC, and Dr. Mohamed G.H. Mohamedali, Mr. Hassen Maatoug, and Mr. Ahmed Soliman from Hezm Mebarek General Hospital-HMC for developing disposable racks for samples transportation, tubes labelling, blood collection and handling. We thank the support provided by Qatar University Biomedical Research Centre, Biosafety Level 3 and Associate Professor Hadi M. Yassine (M.Sc., Ph.D.). We also acknowledge the help by the Anti-Doping Lab-Qatar (ADLQ) and Qatar Red Crescent (QRC) for recruiting control samples. This work was supported by a grant fund from Hamad Medical Corporation (fund number MRC-05-003) and core funding from Qatar Biomedical Research Institute (QBRI).

References

1. International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC). COVID-19 Report: 03 September 2020 [Internet]. https://media.tghn.org/medialibrary/2020/09/ISARIC_Data_Platform_COVID-19_Report_20AUG20.pdf. Published 2020. Accessed.
2. Shi L, Westerhuis JA, Rosen J, Landberg R, Brunius C. Variable selection and validation in multivariate modelling. *Bioinformatics*. 2019;35(6):972-980.
3. Filbin MR, Mehta A, Schneider AM, et al. Plasma proteomics reveals tissue-specific cell death and mediators of cell-cell interactions in severe COVID-19 patients. *bioRxiv*. 2020.
4. Omrani AS, Almaslamani MA, Daghfal J, et al. The first consecutive 5000 patients with Coronavirus Disease 2019 from Qatar; a nation-wide cohort study. *BMC Infect Dis*. 2020;20(1):777.
5. Ramlall V, Thangaraj PM, Meydan C, et al. Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. *Nat Med*. 2020.
6. Meizlish ML, Pine AB, Bishai JD, et al. A neutrophil activation signature predicts critical illness and mortality in COVID-19. *medRxiv*. 2020:2020.2009.2001.20183897.
7. Fraser DD, Cepinskas G, Patterson EK, et al. Novel Outcome Biomarkers Identified With Targeted Proteomic Analyses of Plasma From Critically Ill Coronavirus Disease 2019 Patients. *Crit Care Explor*. 2020;2(9):e0189.
8. Messner CB, Demichev V, Wendisch D, et al. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst*. 2020;11(1):11-24 e14.

9. Shen B, Yi X, Sun Y, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell*. 2020;182(1):59-72 e15.
10. Cava C, Bertoli G, Castiglioni I. In Silico Discovery of Candidate Drugs against Covid-19. *Viruses*. 2020;12(4).
11. Alamdari DH, Moghaddam AB, Amini S, et al. Application of methylene blue -vitamin C -N-acetyl cysteine for treatment of critically ill COVID-19 patients, report of a phase-I clinical trial. *Eur J Pharmacol*. 2020;885:173494.
12. Chen C, Qi F, Shi K, et al. Thalidomide combined with low-dose short-term glucocorticoid in the treatment of critical Coronavirus Disease 2019. *Clin Transl Med*. 2020.
13. Khalil A, Kamar A, Nemer G. Thalidomide-Revisited: Are COVID-19 Patients Going to Be the Latest Victims of Yet Another Theoretical Drug-Repurposing? *Front Immunol*. 2020;11:1248.
14. Caruso A, Caccuri F, Bugatti A, et al. Methotrexate inhibits SARS-CoV-2 virus replication "in vitro". *J Med Virol*. 2020.
15. Yousaf A, Gayam S, Feldman S, Zinn Z, Kolodney M. Clinical Outcomes of CoVid-19 in Patients Taking Tumor Necrosis Factor Inhibitors and/or Methotrexate: A Multi-Center Research Network Study. *J Am Acad Dermatol*. 2020.
16. Hung IF, Lung KC, Tso EY, et al. Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. *Lancet*. 2020;395(10238):1695-1704.
17. Assarsson E, Lundberg M, Holmquist G, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*. 2014;9(4):e95192.
18. Ge SX, Son EW, Yao R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*. 2018;19(1):534.
19. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613.
20. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504.
21. Cotto KC, Wagner AH, Feng YY, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res*. 2018;46(D1):D1068-D1073.
22. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005;21(16):3448-3449.

Table

Table 1: Characteristics of Patients with COVID19 and controls

Variables	Controls (N=50)	Mild-Moderate (N=50)	Severe (N=50)	Total (N=150)	p-value
Age (years)					
Mean \pm SD	37.4 \pm 7.7	40.0 \pm 11.9	45.9 \pm 11.2	41.1 \pm 0.9	<0.001
Median [IQR]	38[33-42]	40[32-51]	47[35-55] *	40[34-49]	
Sex n (%)					
F	2(4.0)	8(16.0)	1(2.0)	11(7.3)	0.015
M	48(96.0)	42(84.0)	49(98.0)	139(92.7)	
Ethnicity n (%)					
Indian subcontinent	43(86.0)	30(60.0)	33(66.0)	106(70.7)	0.035
Middle East North Africa	5(10.0)	15(30.0)	10(20.0)	30(20.0)	
Others	2(4.0)	5(10.0)	7(14.0)	14(9.3)	
BMI (kg/m²)					
Mean \pm SD	25.4 \pm 4.0	26.5 \pm 3.9	29.7 \pm 6.1	27.2 \pm 0.4	<0.001
Median [IQR]	24[23-27]	26[23-28]	28[26-33] * [‡]	26[24-29]	
Obesity Level n (%)					
Normal (\leq 25)	28(56.0)	20(40.0)	7(14.0)	55(36.7)	<0.001
Overweight (25 - 30)	17(34.0)	21(42.0)	25(50.0)	63(42.0)	
Obese (30+)	5(10.0)	9(18.0)	18(36.0)	32(21.3)	
Heart Rate (beats per minute)					
Mean \pm SD	76.5 \pm 10.0	89.8 \pm 16.1	102.4 \pm 16.3	89.3 \pm 1.5	<0.001
Median [IQR]	78[70-82]	86[78-104] [‡]	100[88-117] * [‡]	86[78-100]	
SpO₂ (%)					
Mean \pm SD	98.7 \pm 0.8	98.2 \pm 2.1	93.6 \pm 6.9	79.2 \pm 0.8	<0.001
Median [IQR]	99[98-99]	99[97-100] [‡]	96[91-97] * [‡]	78[72-86]	
SBP (mmHg)					
Mean \pm SD	122.5 \pm 10.3	131.1 \pm 17.7	125.5 \pm 17.6	126.4 \pm 1.3	0.022
Median [IQR]	121[116-130]	129[119-139] *	128[109-137]	126[116-135]	
DBP (mmHg)					
Mean \pm SD	77.9 \pm 7.4	80.2 \pm 9.1	79.5 \pm 12.6	79.2 \pm 0.8	0.503
Median [IQR]	78[73-81]	81[74-88]	78[71-86]	78[72-86]	

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure

* Significantly different than controls

[‡] Significantly different than Mild-Moderate