



Original Research

Analysis of inflammatory protein profiles in the circulation of COVID-19 patients identifies patients with severe disease phenotypes

Nick Keur^{a,1}, Maria Saridaki^{b,1}, Isis Ricaño-Ponce^a, Mihai G. Netea^{a,c,2}, Evangelos J. Giamarellos-Bourboulis^{b,2}, Vinod Kumar^{a,d,e,2,*}

^a Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands

^b 4th Department of Internal Medicine, National and Kapodistrian University of Athens, Athens, Greece

^c Human Genomics Laboratory, Craiova University of Medicine and Pharmacy, Craiova, Romania

^d University of Groningen, University Medical Center Groningen, Department of Genetics, the Netherlands

^e Nitte (Deemed to Be University), Nitte University Centre for Science Education and Research (NUCSER), Deralakatte, Mangalore, India

ARTICLE INFO

Keywords:
Proteomics
COVID-19
Biomarker
Cytokine
Clustering
Endotypes

ABSTRACT

Background: The coronavirus disease (COVID-19) caused by the severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) can present with a broad range of clinical manifestations, ranging from asymptomatic to severe multiple organ failure. The severity of the disease can vary depending on factors such as age, sex, ethnicity, and pre-existing medical conditions. Despite multiple efforts to identify reliable prognostic factors and biomarkers, the predictive capacity of these markers for clinical outcomes remains poor. Circulating proteins, which reflect the active mechanisms in an individual, can be easily measured in clinical practice and therefore may be useful as biomarkers for COVID-19 disease severity. In this study, we sought to identify protein biomarkers and endotypes for COVID-19 severity and evaluate their reproducibility in an independent cohort.

Methods: We investigated a cohort of 153 Greek patients with confirmed SARS-CoV-2 infection in which plasma protein levels were measured using the Olink Explore 1536 panel, which consists of 1472 proteins. We compared the protein profiles from severe and moderate COVID-19 patients to identify proteins associated with disease severity. To evaluate the reproducibility of our findings, we compared the protein profiles of 174 patients with comparable COVID-19 severities in a US COVID-19 cohort to identify proteins consistently correlated with COVID-19 severity in both groups.

Results: We identified 218 differentially regulated proteins associated with severity, 20 proteins were also replicated in an external cohort which we used for validation. Moreover, we performed unsupervised clustering of patients based on 97 proteins with the highest log2 fold changes in order to identify COVID-19 endotypes. Clustering of patients based on differentially regulated proteins revealed the presence of three clinical endotypes. While endotypes 2 and 3 were enriched for severe COVID-19 patients, endotype 3 represented the most severe form of the disease.

Conclusions: These results suggest that identified circulating proteins may be useful for identifying COVID-19 patients with worse outcomes, and this potential utility may extend to other populations.

Trial registration: NCT04357366.

1. Introduction

Coronavirus disease 2019 (COVID-19 disease) caused by severe

acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) induces a variable spectrum of clinical severity: from asymptomatic to mild respiratory symptoms, and to severe cases of multiple organ failure [1,2].

* Corresponding author. Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands.

E-mail addresses: nick.keur@radboudumc.nl (N. Keur), onchocercavol@gmail.com (M. Saridaki), Isis.RicanoPonce@radboudumc.nl (I. Ricaño-Ponce), Mihai.Netea@radboudumc.nl (M.G. Netea), egiamarel@med.uoa.gr (E.J. Giamarellos-Bourboulis), V.Kumar@radboudumc.nl (V. Kumar).

¹ These authors have contributed equally to this work and share first authorship.

² These authors have contributed equally to this work and share senior authorship.

<https://doi.org/10.1016/j.rmed.2023.107331>

Received 13 February 2023; Received in revised form 14 June 2023; Accepted 15 June 2023

Available online 25 June 2023

0954-6111/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

The severity of the disease varies with age, sex, ethnicity, and predisposing comorbidities including obesity and diabetes [3]. Previous studies have also shown differences in severity based on country of origin [4,5]. Although there has been a tremendous effort in identifying reliable prognostic factors and biomarkers, the predictive capacity of such biomarkers for clinical outcomes across a wide range of patient populations is poor.

Circulating biomarkers are the preferred indicators in clinical decision-making to evaluate not only the severity of human diseases but also the effectiveness of treatments. In addition, biological fluids (e.g., blood) can be collected easily and non-invasively. These advantages led to biomarker studies in the serum or plasma of COVID-19 patients. Particularly, proteomic analyses of biological fluids from patients have identified the different patterns of host immune response to COVID-19 [6–8]. These studies have greatly helped to unravel the important immunological mechanisms influencing the disease severity [6,9,10]. However, to what extent the biomarkers identified in one center or patient population is applicable in another patient cohort is still unclear.

In the past years, a technology called multiplex proximity extension assay (PEA) from Olink Proteomics AB has allowed us to simultaneously measure the abundance of multiple proteins. This technology has been used for identifying biomarkers for COVID-19 susceptibility [11,12] and severity [6–8]. For example, by focusing on specific protein panels in a Dutch cohort, we compared the protein profiles of patients admitted to the Intensive Care Units (ICU) and patients not admitted to ICUs (non-ICU) [13]. In that study, 269 proteins were measured using the Olink inflammation I, cardiovascular II, and cardiometabolic panels in the plasma of 153 COVID-19 patients. The study identified 27 proteins differentially regulated between ICU and non-ICU patients. Only a few studies have used the more recent Explore panel, which consists of 1472 proteins.

The biggest study on COVID-19 using the Olink Explore panel was from a cohort of US patients recruited at the Massachusetts General Hospital (MGH) in Boston [6]. It included 384 individuals, of which 80% tested positive for COVID-19. It was a longitudinal study and proteins were measured on the day of the diagnosis with follow-up samples on day 3 and day 7 for the COVID-19 positive patients. The availability of samples from COVID-19 negative patients and the different range in severity within the patients allowed the authors to identify proteins associated not only with COVID-19 susceptibility but also with disease severity.

To test whether protein biomarkers identified in the US COVID-19 cohort are applicable also in a European population we measured the Explore panel in the plasma of COVID-19 patients from Greece, who were part of the SAVE clinical trial (NCT04357366). We first identified proteins that were differentially regulated between patients with severe COVID-19 and patients with moderate COVID-19. Then, we compared our results with those from the US COVID-19 cohort and identified proteins consistently correlated with COVID-19 severity in both cohorts. Clustering the patients based on a subset from the differentially regulated proteins revealed the presence of three different clinical endotypes in this population. We validated the model in the US COVID-19 cohort identifying three clinical endotypes with similar clinical characteristics as in our cohort.

2. Methods

2.1. Cohort description

All patients were part of the open-label non-randomized clinical trial SAVE performed in six sites in Greece. Enrolled patients were adults hospitalized with confirmed infection by SARS-CoV-2 virus by real-time PCR reaction, radiological findings compatible with lower respiratory tract infections; and plasma suPAR level $\geq 6 \mu\text{g/L}$. A detailed description of the cohort has been published by Kyriazopoulou et al. [14] For this study patient severity was determined based on the World Health

Organization Clinical progression scale (WHO scores) [15]. Hospitalized patients requiring oxygen were classified under WHO score 5, while hospitalized patients without oxygen were classified under WHO score 4.

2.2. Olink proteomics

Plasma protein concentrations were analyzed using the Olink Explore 1536 panel, which consists of 1472 unique proteins, covering proteins with broad applicability in neurology, oncology, cardiometabolic, and inflammation. Processing and quality assessment of proteomics data were performed using the “Olink NPX manager” software. The data was transformed and normalized to Olink’s NPX value, NPX is a relative protein quantification unit on a log2 scale, where a difference of 1 NPX equates to a doubling of protein concentration. In addition, three proteins (IL6, TNF, CXCL8) were measured in multiple panels, we removed duplicates and selected one measurement for our data analysis. Further downstream data processing was performed using R (Version 4.1) (R Core Team, 2014). This included the removal of 183 proteins that failed to quantify in more than 75% of samples, or if more than 75% of the protein NPX values fell below the protein limit of detection (LOD) value. The NPX values below the LOD were substituted by the protein’s LOD.

2.3. Statistical analysis

All statistical analyses were performed using R version 4.1. Demographics and other clinical parameters are summarized using descriptive statistics of relevant characteristics. We described continuous variables using the median and inter-quartile range (IQR), whereas for categorical variables frequencies and percentages were used. Significance testing for ordinal or continuous data was performed using the Wilcoxon’s rank sum test (also known as the Mann-Whitney *U* test), whereas the Fisher’s exact test or Chi-square testing were performed for binary data. The non-parametric Spearman’s rank-order correlation was used as a measure for correlation. Analysis of clinical and demographic variables in each identified cluster was performed using the Kruskal–Wallis or Chi-square tests. Figures such as boxplots and Venn diagrams were generated using the ggplot2 (version 2.3.3.5) and ggVenn (version 0.1.9) packages, respectively.

2.4. Differential expression analysis

Differential Expression (DE) analysis was performed using the R package “OlinkAnalyze” (version 1.2.4) provided by Olink, where linear models were fitted for each protein. Models included the WHO score as the main factor with protein NPX values as the dependent variables. Sex, age, and the presence of comorbidities (chronic liver disease, lung disease, heart disease, kidney disease) were included, when possible, as covariables. A model including sex as the main factor and age as a covariate was fitted to identify the protein differentially regulated in both sexes. False discovery rate (FDR) was applied to correct for multiple testing by using the Benjamin-Hochberg method. Proteins were considered significant if the adjusted p-value was <0.05 . Differentially expressed proteins (DEP) were visualized using the package Enhanced-Volcano (version 1.11.5) in R.

For an inter-cohort comparison between our cohort and a previously published study on COVID-19 severity assessing the same proteins [6], we used publicly available Olink proteomic data provided by them. We extracted data from Day 0 ($n = 196$) and acuity levels 3 and 4 (Acuity 3 = 151, Acuity 4 = 45), which are comparable to the WHO scores in the current study. To make this analysis comparable we performed both DE analyses with severity score as the main effect and included age, chronic liver disease, lung disease, heart disease, and kidney disease as covariables, as these variables were available for both studies. To evaluate the proteins that were significant in both studies, significant proteins from

each study were extracted and intersected using unique protein IDs, as three proteins were measured in multiple panels.

2.5. Unsupervised clustering and UMAP projection

Unsupervised clustering was performed using hierarchical clustering and the dimensional reduction method Uniform Manifold Approximation and Projection (UMAP). We included proteins that reached the threshold of adjusted p-value <0.05 and log2 fold change ± 0.40 . Thereafter, for each protein, NPX values were mean-centered and scaled to have a standard deviation of 1. The heatmap was generated using the package ComplexHeatmap (version 2.10.0) [16].

2.6. Gene set enrichment and pathway analysis

Pathway and network enrichment was performed using the R package PathfindR (version 1.1.2) [17]. This tool was utilized to identify enriched pathways through active subnetworks searching using protein-protein interaction networks. The default parameters were selected, which are 10 iterations; Protein-protein interaction: Biogrid; p-values adjustment: “Bonferroni,” adjusted p-value threshold: <0.05). Databases used for this analysis were KEGG, Reactome pathways, and GO terms of biological processes, molecular function, and cellular component. Both the bubble plot and heatmaps are generated using ggplot2.

3. Results

3.1. Proteomic profiling of COVID-19 patients

Our cohort comprised 153 Greek patients with confirmed SARS-CoV-2 infection. The median age of patients was 60 years (IQR, 51–73 years) and 64% (N = 98) were male. Patient severity was determined using the WHO scores. Most of the patients (68.7%, N = 105) had severe disease (WHO score 5), while the remainder 31.3% of the patients (N = 48) had moderate disease (WHO score 4). A detailed overview of the clinical and demographic characteristics of the cohort can be found in Table 1.

To assess the differences in the proteome profile of COVID-19 patients in plasma, targeted proteomic profiling was performed using the Olink Explore 1538/384 panel. After quality assessment and data processing, 1280 proteins were available for comparative analysis. Since both, age, and sex are considered strong risk factors for COVID-19 severity and outcomes, we systematically tested for these effects by projecting the protein expression patterns of the COVID-19 samples in a 2-dimensional space using UMAP (Additional file 1, Supplementary Fig. 1A). We did not observe any differences in the proteomic profile based either on sex or age using this approach. However, by performing a differential abundance analysis we found that the concentration of ten

proteins significantly differs (adjusted P < 0.05) between males and females (Additional file 1: Supplementary Fig. 1B, Additional file 2: Supplementary Table 1). Moreover, the abundance of 327 proteins correlated with age (Additional file 2: Supplementary Table 2). The top five proteins whose levels increase the most with age are EDA2R (r = 0.61), FSTL3 (r = 0.52), IGFBPL1 (r = 0.53), NEFL (r = 0.52), and REG4 (r = 0.52), whereas the top five proteins that had decreased concentrations with age were EGFR (r = -0.31), FETUB (r = -0.33), IGFBP3 (r = -0.37), NELL1 (r = -0.39), UMOD (r = -0.33).

3.2. Proteins associated with COVID-19 severity

We subsequently identified proteins associated with COVID-19 severity in the plasma by comparing the proteomic profile of patients with WHO-scores 4 and 5, while computationally adjusting for sex, age, and the presence of comorbidities (chronic liver disease, lung disease, heart disease, kidney disease). We identified 218 proteins that were significantly associated with COVID-19 severity, of which 131 displayed higher concentrations in patients with WHO-score 5 and 87 displayed lower concentrations (Fig. 1A, Additional file 2: Supplementary Table 3). The most significantly increased proteins in severe COVID-19 patients were Keratin 19 (KRT19) related to keratinization and epithelial cell injury, Interleukin 1 Receptor Like 1 (IL1RL1), TNF Receptor Superfamily Member 10b (TNFRSF10B) involved in apoptosis, and V-Set and Immunoglobulin Domain Containing 4 (VSG4) a negative regulator of T-cell proliferation. In contrast, the most significant proteins that had lower concentrations in patients with more severe COVID-19 (Fig. 1B) were carbonic anhydrase 6 (CA6) involved in carbonate dehydratase activity, Cell adhesion molecule-related/down-regulated by oncogenes (CDON) belonging to the immunoglobulin superfamily of cell-adhesion molecules, Kit ligand (KITLG), and Fas Ligand (FASLG) involved in apoptosis. Interestingly, the significant dysregulated proteins were distributed in the four different categories of the inflammation panel: 23% belong to neurology, 23% to oncology, 24% to cardiometabolic, and 30% to inflammation, suggesting the interplay between multiple pathways in COVID-19 pathophysiology (Additional file 1: Supplementary Fig. 4).

3.3. Pathway enrichment analyses on proteins associated with severity

To obtain a general understanding of the functional implications of the differentially regulated proteins, we performed functional pathway and network enrichment analyses using the 218 proteins significantly dysregulated with COVID-19 severity. We observed many pathways related to cytokine regulation response and other immune-related functions using different databases (Fig. 1C, Additional file 2: Supplementary Tables 4–8). Using the KEGG database, 123 pathways were significant (Additional file 2: Supplementary Table 4), including cytokine-cytokine receptor interaction, pathogenic *Escherichia coli* infection, necroptosis, apoptosis, and Influenza A. Based on the Reactome database 106 pathways were significantly enriched (Additional file 2: Supplementary Table 5), including regulation of necroptotic cell death, TNFs bind their physiological receptors, extracellular matrix organization, signal transduction by L1, TNF receptor superfamily (TNFSF) members mediating non-canonical NF- κ B pathway and post-translational protein phosphorylation. We also identified 47 pathways based on GO biological terms (Additional file 2: Supplementary Table 6), 15 based on GO biological function (Additional file 2: Supplementary Table 7), and 26 based on GO cellular function (Additional file 2: Supplementary Table 8). The most significant pathways across databases are shown in Fig. 1C.

3.4. Shared circulatory proteins between US and Greek COVID-19 cohorts

To validate the associations of the proteins identified in this study

Table 1

A detailed overview of the clinical and demographic characteristics of the cohort.

Characteristic	Overall, N = 153 ^a	WHO score 4, N = 48 ^a	WHO score 5, N = 105 ^a	p-value ^b
Age	60 (51, 73)	57 (48, 71)	63 (52, 74)	0.1
Sex				0.5
Female	55 (36%)	19 (40%)	36 (34%)	
Male	98 (64%)	29 (60%)	69 (66%)	
Diabetes mellitus	39 (25%)	12 (25%)	27 (26%)	>0.9
Congestive heart failure	12 (7.8%)	3 (6.2%)	9 (8.6%)	0.8
Chronic kidney disease	1 (0.7%)	0 (0%)	1 (1%)	>0.9
Chronic heart disease	15 (9.8%)	4 (8.3%)	11 (10%)	0.8

^a Median (IQR); n (%).

^b Fisher's Exact Test; Chi-squared Test; Wilcoxon rank sum Test.

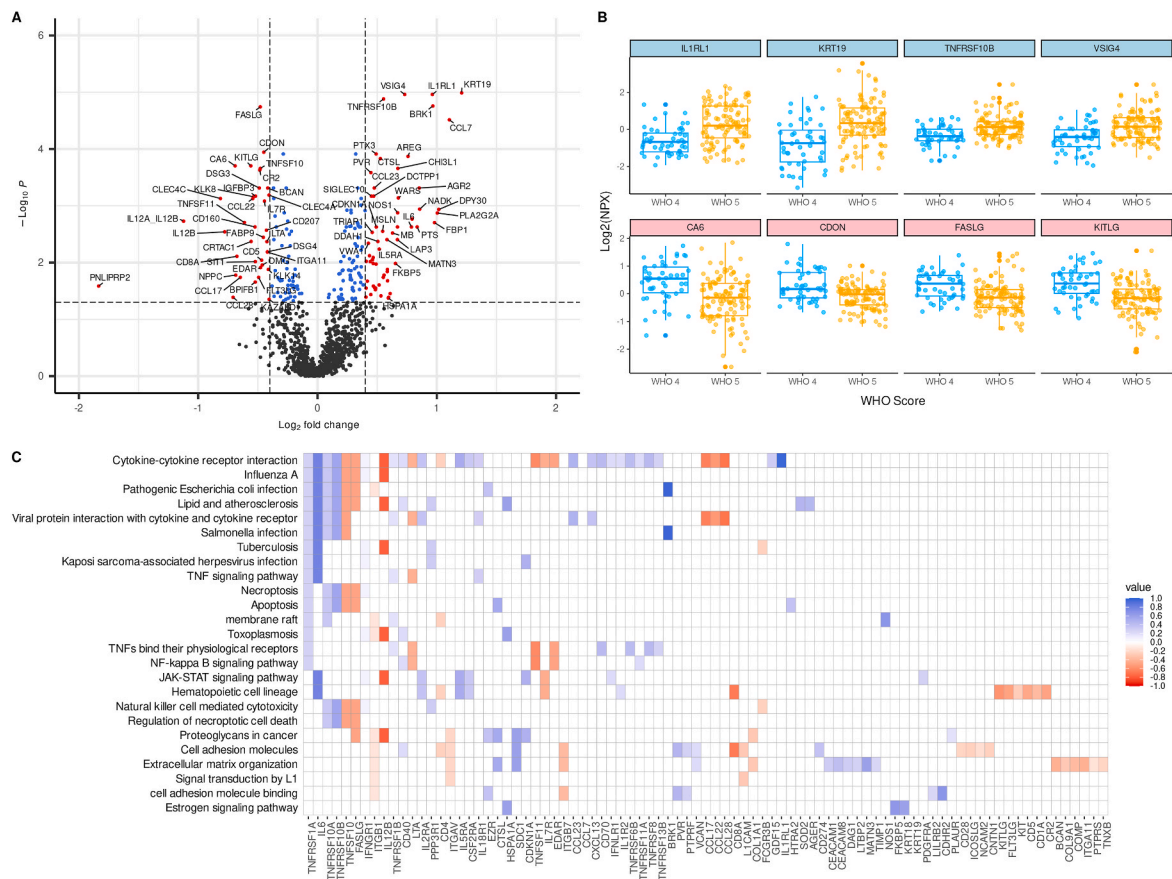


Fig. 1. Identification of proteins associated with severity in the SAVE cohort. (a) The volcano plot shows the proteins associated with severity in the SAVE cohort. Linear models are fitted for each protein using the WHO severity score (WHO score 4 = 48, WHO score 5 = 105). Proteins with a positive log2 fold change indicate that an increase in protein level is associated with increased disease severity and proteins with a negative log2 fold change indicate decreasing protein levels with severity. P-values are adjusted using the Bonferroni adjustment. Proteins colored in red have both, a log2 fold change $> \pm 0.4$ and adjusted p-value < 0.05 , whereas proteins colored in blue have a log2 fold change $< \pm 0.4$ and are below the $< \text{adjusted p-value } 0.05$, finally proteins in grey are non-significant. (b) Boxplot shows the four most significant proteins in both directions. (blue = Up-regulated, red = Down-regulated). The x-axis shows the WHO score for each individual. (c) Heatmap visualizing enriched terms and the associated proteins. We selected proteins that both have a log2 fold change $> \pm 0.4$ and adjusted p-value < 0.05 . Rows represent enriched terms, whereas the columns represent proteins.

with COVID-19 severity in an independent cohort, we used the public Olink proteomic data provided by Filbin et al. [6]. This US COVID-19 cohort comprised 306 confirmed COVID-19 patients and 78 COVID-19 negative patients. Proteomic profiling was performed on days 0, 3, and 7 for the COVID-19 positive patients, while COVID-19 negative patients had a single sample taken on day 0 and served as a control group. Thereafter, patients were classified into five groups (A1 - died, A2 - intubated and survived, A3 - hospitalized on oxygen; A4 - hospitalized without oxygen, A5 - discharged) according to the disease severity. For this inter-cohort analysis, we selected the patients with acuity levels equivalent to our definition of severity on day 0 (acuity levels 3 and 4, $N = 133$ and $N = 41$, respectively). Next, we performed differential expression analysis and identified 46 proteins differently regulated between acuity levels 3 and 4. However, this included IL6 which was measured in multiple panels. After removing the duplicated measurements, 31 unique proteins were differentially regulated. Twenty-seven proteins were upregulated with disease severity (Fig. 2A, Additional file 2: Supplementary Table 9), while 4 proteins were down-regulated (BID, CA6, CDON, and TNFSF11). The top 4 up- and down-regulated proteins are shown in Fig. 2B. These proteins were enriched in pathways related to viral protein interaction with cytokine and cytokine receptor, and the chemokine signaling pathway ($p = 4.09 \times 10^{-6}$ and $p = 1.49 \times 10^{-5}$, respectively). Out of the 31 differentially regulated proteins in this cohort, 20 proteins (Fig. 2C) were also replicated in our cohort with similar log2 fold changes. (Fig. 2D). From the

11 proteins that failed to replicate, two were excluded due to quality control in our study (GOLM2 and JUN), and thus were not tested.

3.5. Identification of COVID-19 endotypes in European patients

Next, given the heterogeneous nature of the COVID-19 phenotypes, we tested whether based on protein profiles it is possible to identify subgroups (endotypes) of COVID-19 patients in the SAVE cohort. Therefore, we performed unsupervised hierarchical clustering using 218 significantly differentially expressed proteins with COVID-19 severity. This analysis identified three clearly defined clusters of COVID-19 patients (Additional file 1: Supplementary Fig. 2). We assessed if we could identify the same clusters using 97 proteins based on the expression levels of the most significantly differentially regulated proteins ($\log_2FC > \pm 0.4$ and adjusted p-value < 0.05), and we were able to confirm it (Fig. 3A, Additional file 1: Supplementary Fig. 3). To further characterize the clusters, we compared different demographic and clinical parameters. No significant differences in sex, age, comorbidity index, or patient state index were observed between the clusters (Additional file 1: Supplementary Table 1). However, significant enrichment was observed using the WHO score on day 1, particularly cluster 1 is enriched for patients with a WHO score of 4 (80%, $N = 28$), while cluster 2 and cluster 3 are enriched for a WHO score of 5 (78%, $N = 58$ and 91%, $N = 41$, respectively). Moreover, we also observed significant differences among the clusters in the concentrations of soluble urokinase

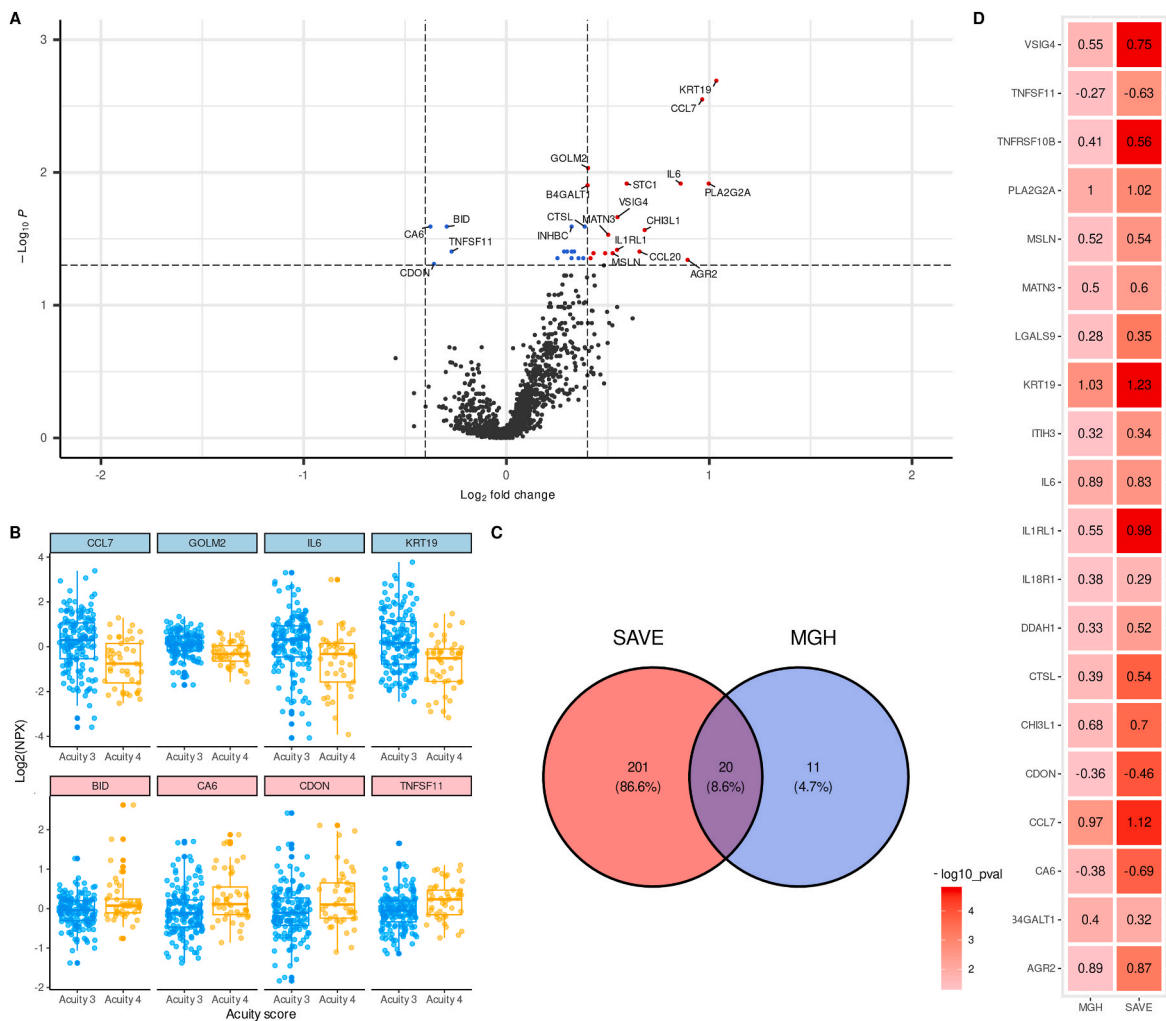


Fig. 2. Inter-cohort identification of proteins associated with severity. (a) Volcano plot visualizing proteins associated with acuity score in MGH cohort. For each protein, a linear model is fitted using the Acuity score from day 0. Positive values for log₂ fold changes indicate increasing protein levels with increased severity, while negative values mean the opposite. (b) Boxplot shows the four most significant proteins associated with Acuity levels in both directions. (c) Venn-diagram visualizing overlapping significant associations proteins from each differential abundance analysis between our own cohort of Covid-19 patients and the MGH Covid-19 cohort including age, chronic liver disease, lung disease, heart disease, and kidney disease as covariables. For both analyses, we used the Bonferroni correction to correct for multiple testing and used 0.05 as the significance threshold. (d) Heatmap visualization which shows the effect size and p-values for the overlapping proteins.

plasminogen activator receptor (suPAR, p-value = 0.0006), Apache (p-value = 0.0002), and SOFA (p-value <0.0001) scores indicating different degrees of disease severity and organ damage. Patients in cluster 1 have a milder form of the disease, characterized by lower suPAR concentrations, Apache, and sofa scores compared to clusters 2 and 3. We did not observe significant differences while comparing clusters 2 and 3.

3.6. Comparing cluster 1 and cluster 3 endotypes identified specific pathways associated with COVID-19 severity

To assess the difference in the proteomic profile of the three endotypes, we performed differential expression analysis using the identified endotypes as our main factor. The comparison between clusters 1 and 3 revealed that 755 proteins were differentially regulated after correcting for multiple-hypothesis-testing (Additional file 2: [Supplementary Table 10](#)). Most of the proteins were up-regulated in cluster 3 compared to cluster 1 and only 151 proteins were down-regulated. A similar comparison between clusters 1 and 2 identified 303 proteins differentially regulated, with 61 proteins down-regulated in cluster 2 and 242 up-regulated (Additional file 2: [Supplementary Table 11](#)). Finally, 513

proteins were differentially regulated while comparing clusters 2 and 3, most of the proteins were up-regulated in cluster 3, and only 70 proteins were down-regulated (Additional file 2: [Supplementary Table 12](#)).

Pathway analysis on the differentially expressed proteins from clusters 1 and 3 (Fig. 3B, Additional file 2: [Supplementary Table 13](#)) based on the KEGG database revealed 163 significant pathways. Most of them (118) overlapped with the pathways from the analysis based on the WHO score. However, 45 pathways were specific for the cluster analysis, the 10 most significant specific pathways were: peroxisome, RIG-I-like receptor signaling pathway, IL-17 signaling pathway, cytosolic DNA-sensing pathway, base excision repair, proteasome, VEGF signaling pathway, chemokine signaling pathway and amoebiasis. These results suggest that the dysregulation of these pathways is linked to organ damage and increased severity.

3.7. Validation of identified endotypes using an external validation cohort

To validate the model for the identification of the endotypes in an independent cohort, we performed an unsupervised hierarchical clustering in the US cohort from Filbin et al. [6] using the 97 dysregulated proteins associated with COVID-19 severity in our cohort. In

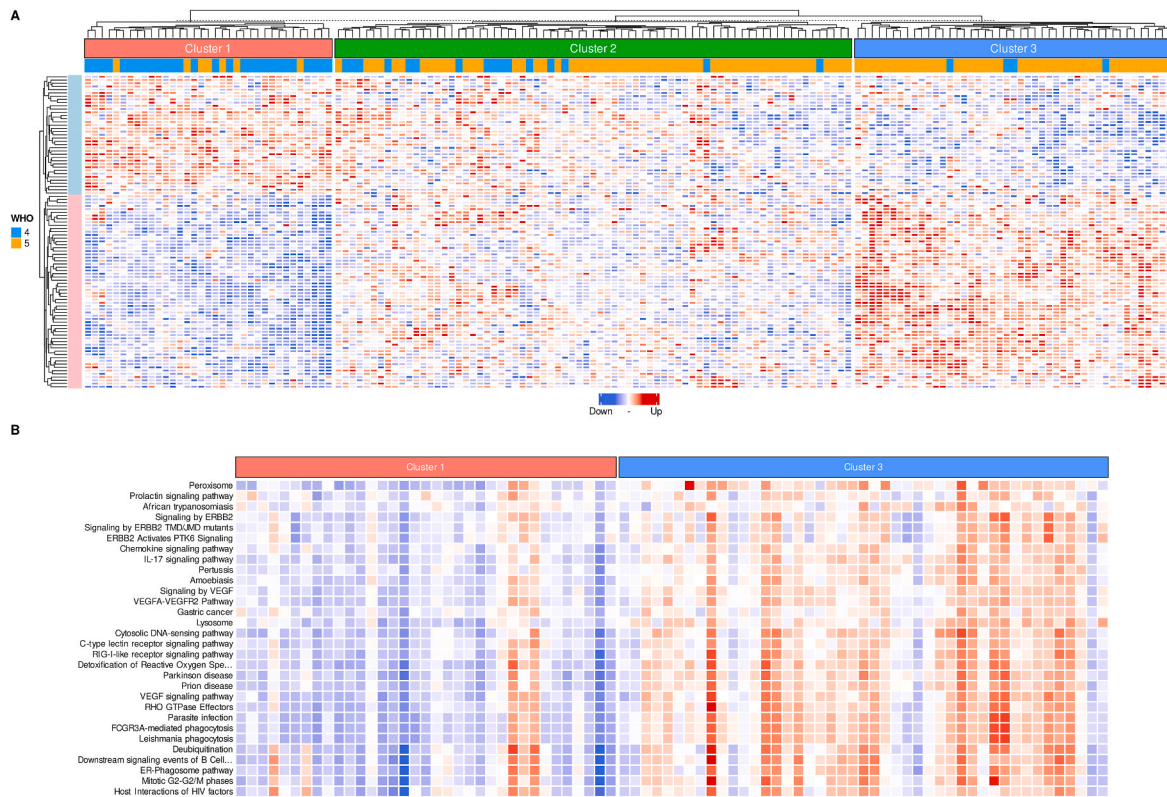


Fig. 3. Heatmap with DE proteins and pathways associated with severity (A) Heatmap visualizing proteins associated with severity. We selected all proteins that were found to be significantly associated with severity after multiple testing corrections. ($\log_2FC \geq \pm 0.4$ and adjusted p-value < 0.05). Rows represent proteins, whereas the columns represent individuals. Proteins and individuals are both ordered by hierarchical clustering. Top-annotation shows the WHO score for each individual. (B) Heatmap shows the comparison of selected terms in individual samples between Cluster 1 and Cluster 3. The y-axis represents the enriched pathway term, whereas the x-axis represents individual samples. The color indicates the aggregated z-score of each enriched term per sample, red indicates an overall increased expression (activated) of the corresponding enriched term, whereas the blue color indicates decreased expression (repressed).

concordance with our study, we observed three clusters of COVID-19 patients (Additional file 1: [Supplementary Fig. 3](#)). The characterization of clusters was performed with available demographic and clinical parameters. Similar as before, we observed no differences in age, BMI, and occurrences of comorbidities such as lung disease, kidney disease, and diabetes between the clusters, except for heart disease (P-value = 0.042, Additional file 1: [Supplementary Table 2](#)). However, we observed significant differences in severity using the Acuity score (p-value < 0.004) indicating different degrees of severity within the specified spectrum. Additional information such as immune cell counts (lymphocytes neutrophils and monocytes), creatine, C-reactive protein (CRP), d-dimer and lactate dehydrogenase (LDH) were also available. We observed consistent upregulation of these clinical parameters in cluster 3 (Additional file 1: [Supplementary Table 3](#)) except for monocyte numbers which were not different between the clusters.

4. Discussion

In our study, we used a comprehensive proteomics approach to measure a large number of proteins in the plasma of COVID-19 patients with moderate and severe disease. Our study reports, in addition to large proteomics data from COVID-19 patients, several important findings.

First, age and sex have been reported as important factors predisposing to COVID-19 severity [18–22] and our results are in concordance with the reports that age is an important determinant of COVID-19 severity. Although concentrations of 327 proteins were significantly correlated with age in our cohort, 21% also showed a correlation with disease severity even after correcting for age suggesting their role in influencing disease severity across different adult patient age categories. This includes proteins such as KRT19, VSIGA, TNFRSF10B, and FASL.

Second, we identified 218 proteins that are significantly different between severe and moderate COVID-19 disease. It is important to validate how many of these proteins are also associated with COVID-19 severity in independent studies. In this context, a recent study has compared five different studies that applied the Olink affinity proteomics platform to analyse the proteomic profile between COVID-19 patients and controls [7]. Thirteen proteins, out of 253 tested, were consistently associated with COVID-19 susceptibility in all studies: CCL16, CCL7, CXCL10, CCL8, LGALS9, CXCL11, IL1RN, CCL2, CD274, IL6, IL18, MERTK, IFNG, and IL18R1. The authors hypothesized that the heterogeneity of the studies might be due to the difference in the disease severity of the patients, and the difference in comorbidities in the controls included in the studies. Nevertheless, six of those proteins were also differentially regulated in our study of COVID-19 severity (CCL7, LGALS9, CD274, IL6, MERTK, IL18R1), suggesting that their dysregulation is associated with disease severity.

When we focus on COVID-19 severity comparing studies is even more challenging, as the severity of the disease was classified using different indicators. For example, in the US cohort study [6] the patients were classified based on an acuity level defined by the authors, while we used the WHO clinical progression score. To overcome this difference, in this study, we focused only on groups that matched our established criteria for defining severity (oxygen requirement). Hence, out of the 306 individuals included in the US cohort, we limited our analysis to 174 individuals. This also allowed us to have comparable sample sizes. Moreover, in our cohort, 68.5% of the patients required oxygen, while in the US cohort 76.4%. Even with this comparable sample size and definition of severity, we only replicated 8.6% of the proteins identified in our cohort using the US cohort. It is important to note that we only identified 31 proteins differentially regulated in the US cohort, while

221 were dysregulated in our cohort. Moreover, it is important to mention that our cohort included only Greek individuals, while the US cohort comprised white, black, and Hispanic individuals. In the initial analysis of the US cohort, this was corrected, but the information was not publicly available and could not be included in our analysis. Thus, the contribution of ethnicity should be further explored. These results suggest that COVID-19 patient cohorts are extremely heterogeneous and large-scale population-specific biomarker studies might be helpful to explain this heterogeneity.

Third, our analysis of identifying endotypes within the European cohort identified significantly different protein signatures between patients. The fact that we used a large number of proteins instead of a specific panel with a limited number of proteins provided this resolution to identify subtypes of patients. Patients belonging to cluster 3 in our analysis were characterized by extreme severity, and proteins associated with this cluster were enriched for many pathways including 45 pathways not identified based on the WHO-score. Among these pathways, we observed the complement and coagulation cascades, RIG-I-like receptor signaling, and IL-17 signaling pathway, which have extensively been described in the COVID-19 literature. Interestingly, these pathways were not among the significant pathways using the WHO score suggesting that patient clusters based on the proteome reduce the heterogeneity and therefore increase the sensitivity to detect more specific pathways. Although it still needs to be investigated how proteins involved in pathways regulating pluripotency of stem cells and phosphatidylinositol signaling system determine the incidence of respiratory failure and mortality, it is an important finding.

Finally, we validated the set of proteins for the identification of the Covid-19 endotypes and the endotypes in the US COVID-19 cohort. We identified three clusters with different degrees of severity and levels of clinical parameters indicating organ damage, inflammation, and coagulation problems. These results indicate that the identified endotypes can be extrapolated to other populations and cohorts.

The main strength of this study is that it does not limit to the critically ill patients hospitalized in the ICU, but it is focusing on the clustering of patients hospitalized in the general wards without and with the need for oxygen well before critical illness develops. A study using the Explore panel to investigate the plasma proteome profile of COVID-19 patients with mild to moderate symptoms has been performed by Zhong et al. [8] Their study focused on comparing the protein profile of the patients at the time of diagnosis and 14 days later, identifying 239 differentially regulated proteins. Although the authors focused on a less severe group of patients, in which only 4% of their cohort had breathing issues, they observed that most proteins are elevated or down-regulated in a similar way as in the US COVID-19 cohort which includes severe cases (76.4% required oxygen).

5. Conclusions

In summary, by profiling a large number of plasma proteins our study was able to not only identify consistently associated proteins across different studies, but it also showed the possibility to identify subtypes of inflammatory endotypes of COVID-19 patients. Future work is needed to characterize these endotypes further to understand the mechanistic basis for this heterogeneity.

Ethics approval and consent to participate

Human subjects: Approved in Greece by National Ethics Committee approval 38/20; National Organization for Medicines approval ISO 28/20. Written informed consent was provided by the patient or legal representative before screening. Clinical trial registration EudraCT number 2020-001466-11 and <https://clinicaltrials.gov/registration/NCT04357366>.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are available contained in this published article and its supplementary information files. In addition, the datasets are accessible by the corresponding authors upon reasonable request.

Funding

This work was funded by the ZonMw (10430012010002) to VK. The SAVE study was funded by the Hellenic Institute for the Study of Sepsis, Swedish Orphan Biovitrum, and the Horizon 2020 Framework Programme. (NCT04357366)

Credit author statement

Vinod Kumar, conceived the study, revised the manuscript. All authors read and approved the final manuscript., Evangelos J. Giamarellos-Bourboulis, conceived the study, revised the manuscript. All authors read and approved the final manuscript., Mihai Netea, conceived the study, revised the manuscript. All authors read and approved the final manuscript., Maria Saridaki, acquired the data, Nick Keur, analyzed the data and prepared the figures and tables, drafted the manuscript, All authors read and approved the final manuscript, Isis Ricaño-Ponce, analyzed the data and prepared the figures and tables, drafted the manuscript, All authors read and approved the final manuscript.

Declaration of competing interest

All authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank all the participants who agreed to take part in this study and all the members who contributed to the selection and assessment.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rmed.2023.107331>.

List of abbreviations

COVID-19	Coronavirus-associated disease 2019
DE	Differential expression
ICU	Intensive Care Units
IQR	Inter quartile range
LOD	Limit of detection
PSA	proximity extension assay
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
WHO	World Health Organization

References

- [1] M. Lopes-Pacheco, P.L. Silva, F.F. Cruz, D. Battaglini, C. Robba, P. Pelosi, M. Morales, C. Caruso Neves, P.R.M. Rocco, Pathogenesis of multiple organ injury in COVID-19 and potential therapeutic strategies, *Front. Physiol.* 12 (2021) 29, <https://doi.org/10.3389/FPHYS.2021.593223/BIBTEX>.
- [2] T. Mokhtari, F. Hassani, N. Ghaffari, B. Ebrahimi, A. Yarahmadi, G. Hassanzadeh, COVID-19 and multiorgan failure: a narrative review on potential mechanisms, *J. Mol. Histol.* 51 (2020) 613, <https://doi.org/10.1007/S10735-020-09915-3>.

- [3] A. Magdy Beshbishy, V.B. Oti, D.E. Hussein, I.F. Rehan, O.S. Adeyemi, N. Rivero-Perez, A. Zaragoza-Bastida, M.A. Shah, K. Abouelezz, H.F. Hetta, N. Cruz-Martins, G.E.S. Batiha, Factors behind the higher COVID-19 risk in diabetes: a critical review, *Front. Public Health* 9 (2021) 637, <https://doi.org/10.3389/fpubh.2021.591982/BIBTEX>.
- [4] R. Mathur, C.T. Rentsch, C.E. Morton, W.J. Hulme, A. Schultze, B. MacKenna, R. M. Eggo, K. Bhaskaran, A.Y.S. Wong, E.J. Williamson, H. Forbes, K. Wing, H. I. McDonald, C. Bates, S. Bacon, A.J. Walker, D. Evans, P. Inglesby, A. Mehrkar, H. J. Curtis, N.J. DeVito, R. Croker, H. Drysdale, J. Cockburn, J. Parry, F. Hester, S. Harper, I.J. Douglas, L. Tomlinson, S.J.W. Evans, R. Grieve, D. Harrison, K. Rowan, K. Khunti, N. Chaturvedi, L. Smeeth, B. Goldacre, Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform, *Lancet (London, England)* 397 (2021) 1711, [https://doi.org/10.1016/S0140-6736\(21\)00634-6](https://doi.org/10.1016/S0140-6736(21)00634-6).
- [5] S. Sze, D. Pan, C.R. Nevill, L.J. Gray, C.A. Martin, J. Nazareth, J.S. Minhas, P. Divall, K. Khunti, K.R. Abrams, L.B. Nellums, M. Pareek, Ethnicity and clinical outcomes in COVID-19: a systematic review and meta-analysis, *EclinicalMedicine* 29–30 (2020), 100630, <https://doi.org/10.1016/j.eclim.2020.100630>.
- [6] M.R. Filbin, A. Mehta, A.M. Schneider, K.R. Kays, J.R. Guess, M. Gentili, B. G. Fenyes, N.C. Charland, A.L.K. Gonye, I. Gushterova, H.K. Khanna, T.J. LaSalle, K.M. Lavin-Parsons, B.M. Lilley, C.L. Lodenstein, K. Manakongtreecheep, J. D. Margolin, B.N. McKaig, M. Rojas-Lopez, B.C. Russo, N. Sharma, J. Tantivit, M. F. Thomas, R.E. Gerszten, G.S. Heimberg, P.J. Hoover, D.J. Lieb, B. Lin, D. Ngo, K. Pelka, M. Reyes, C.S. Smillie, A. Waghay, T.E. Wood, A.S. Zajac, L.L. Jennings, I. Grundberg, R.P. Bhattacharyya, B.A. Parry, A.C. Villani, M. Sade-Feldman, N. Hacohen, M.B. Goldberg, Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions, *Cell Reports Med* 2 (2021), 100287, <https://doi.org/10.1016/j.xcrm.2021.100287/ATTACHMENT/D155F083-9918-4A3D-B2FB-A376B50ED864/MMC8.XLSX>.
- [7] K. Suhre, H. Sarwath, R. Engelke, M.U. Sohail, S.J. Cho, W. Whalen, S. Alvarez-Mulet, J. Krumsiek, A.M.K. Choi, F. Schmidt, Identification of robust protein associations with COVID-19 disease based on five clinical studies, *Front. Immunol.* 12 (2022) 5935, <https://doi.org/10.3389/fimmu.2021.781100/BIBTEX>.
- [8] W. Zhong, O. Altay, M. Arif, F. Edfors, L. Doganay, A. Mardinoglu, M. Uhlen, L. Fagerberg, Next generation plasma proteome profiling of COVID-19 patients with mild to moderate symptoms, *EBioMedicine* 74 (2021), 103723, <https://doi.org/10.1016/j.ebiom.2021.103723/ATTACHMENT/BE27A4B1-F415-43C3-8E1B-D202CD8BCC0E/MMC10.XLSX>.
- [9] D. Blanco-Melo, B.E. Nilsson-Payant, W.C. Liu, S. Uhl, D. Hoagland, R. Möller, T. X. Jordan, K. Oishi, M. Panis, D. Sachs, T.T. Wang, R.E. Schwartz, J.K. Lim, R. A. Albrecht, B.R. tenOever, Imbalanced host response to SARS-CoV-2 drives development of COVID-19, *Cell* 181 (2020) 1036–1045.e9, <https://doi.org/10.1016/j.cell.2020.04.026>.
- [10] A. Stukalov, V. Girault, V. Grass, O. Karayel, V. Bergant, C. Urban, D.A. Haas, Y. Huang, L. Oubraham, A. Wang, M.S. Hamad, A. Piras, F.M. Hansen, M.C. Tanzer, I. Paron, L. Zinzula, T. Engleitner, M. Reinecke, T.M. Lavacca, R. Ehmann, R. Wölfel, J. Jores, B. Kuster, U. Protzer, R. Rad, J. Ziebuhr, V. Thiel, P. Scaturro, M. Mann, A. Pichlmair, Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV, *Nat* 594 (2021) 246–252, <https://doi.org/10.1038/s41586-021-03493-4>, 2021 5947862.
- [11] A.B. Palmos, V. Millischer, D.K. Menon, T.R. Nicholson, L.S. Taams, B. Michael, G. Sunderland, M.J. Griffiths, C. Hübel, G. Breen, Proteome-wide Mendelian randomization identifies causal links between blood proteins and severe COVID-19, *PLoS Genet.* 18 (2022), e1010042–e1010042, <https://doi.org/10.1371/JOURNAL.PGEN.1010042>.
- [12] S. Zhou, G. Butler-Laporte, T. Nakanishi, D.R. Morrison, J. Afilalo, M. Afilalo, L. Laurent, M. Pietzner, N. Kerrison, K. Zhao, E. Brunet-Ratnasingham, D. Henry, N. Kimchi, Z. Afrasiabi, N. Rezak, M. Bouab, L. Petitjean, C. Guzman, X. Xue, C. Tselios, B. Vulesevic, O. Adeleye, T. Abdullah, N. Almamlouk, Y. Chen, M. Chassé, M. Durand, C. Paterson, J. Normark, R. Frithiof, M. Lipsey, M. Hultström, C.M.T. Greenwood, H. Zeberg, C. Langenberg, E. Thysell, M. Pollak, V. Mooser, V. Forgetta, D.E. Kaufmann, J.B. Richards, A Neanderthal OAS1 isoform protects individuals of European ancestry against COVID-19 susceptibility and severity, *Nat. Med.* 274 (2021) 659–667, <https://doi.org/10.1038/s41591-021-01281-1>, 27 (2021).
- [13] N.A.F. Janssen, I. Grondman, A.H. de Nooijer, C.K. Boahen, V.A.C.M. Koeken, V. Matzaraki, V. Kumar, X. He, M. Kox, H.J.P.M. Koenen, R.L. Smeets, I. Joosten, R. J.M. Brüggemann, I.J.E. Kouijzer, H.G. van der Hoeven, J.A. Schouten, T. Frenzel, M.H.E. Reijers, W. Hoefsloot, A.S.M. Dofferhoff, M.J. van Apeldoorn, M.J. T. Blaauw, C. Veerman, C. Maas, A.H. Schoneveld, I.E. Hoefler, L.P.G. Derde, M. van Deuren, J.W.M. van der Meer, R. van Crevel, E.J. Giamarellos-Bourboulis, L.A. B. Joosten, M.M. van den Heuvel, J. Hoogerwerf, Q. de Mast, P. Pickkers, M. G. Netea, F.L. van de Veerdonk, Dysregulated innate and adaptive immune responses discriminate disease severity in COVID-19, *J. Infect. Dis.* 223 (2021) 1322–1333, <https://doi.org/10.1093/infdis/jiab065>.
- [14] E. Kyriazopoulou, P. Panagopoulos, S. Metallidis, G.N. Dalekos, G. Poulakou, N. Gatselis, E. Karakike, M. Saridakis, G. Loli, A. Stefanos, D. Prasianaki, S. Georgiadou, O. Tsachouridou, V. Petrakis, K. Tsiakos, M. Kosmidou, V. Lygoura, M. Dareioti, H. Millionis, I.C. Papanikolaou, K. Akinosoglou, D.M. Myrodis, A. Gravani, A. Stamou, T. Gkavogianni, K. Katrini, T. Marantos, I.P. Trontzas, K. Syrgios, M. Chatzis, S. Chatzis, N. Vechlidis, C. Avgoustou, S. Chlantzis, M. Kyrianiou, J.W.M. van der Meer, J. Eugen-Olsen, M.G. Netea, E.J. Giamarellos-Bourboulis, An open label trial of anakinra to prevent respiratory failure in covid-19, *Elife* 10 (2021), <https://doi.org/10.7554/eLife.66125>.
- [15] J.C. Marshall, S. Murthy, J. Diaz, N. Adhikari, D.C. Angus, Y.M. Arabi, K. Baillie, M. Bauer, S. Berry, B. Blackwood, M. Bonten, F. Bozza, F. Brunkhorst, A. Cheng, M. Clarke, V.Q. Dat, M. de Jong, J. Denholm, L. Derde, J. Dunning, X. Feng, T. Fletcher, N. Foster, R. Fowler, N. Gobat, C. Gomersall, A. Gordon, T. Glueck, M. Harhay, C. Hodgson, P. Horby, Y.J. Kim, R. Kojan, B. Kumar, J. Laffey, D. Malvey, I. Martin-Loeches, C. McArthur, D. McAuley, S. McBride, S. McGuinness, L. Merson, S. Morpeth, D. Needham, M. Netea, M.D. Oh, S. Phyu, S. Piva, R. Qiu, H. Salisu-Kabara, L. Shi, N. Shimizu, J. Sinclair, S. Tong, A. Turgeon, T. Uyeki, F. van de Veerdonk, S. Webb, P. Williamson, T. Wolf, J. Zhang, A minimal common outcome measure set for COVID-19 clinical research, *Lancet Infect. Dis.* 20 (2020), [https://doi.org/10.1016/S1473-3099\(20\)30483-7](https://doi.org/10.1016/S1473-3099(20)30483-7) e192–e197.
- [16] Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, *Bioinformatics* 32 (2016) 2847–2849, <https://doi.org/10.1093/BIOINFORMATICS/BTW313>.
- [17] E. Ulgen, O. Ozisik, O.U. Sezerman, PathfindR: an R package for comprehensive identification of enriched pathways in omics data through active subnetworks, *Front. Genet.* 10 (2019) 858, <https://doi.org/10.3389/fgene.2019.00858/BIBTEX>.
- [18] Y. Li, M. Jerkic, A.S. Slutsky, H. Zhang, Molecular mechanisms of sex bias differences in COVID-19 mortality, *Crit. Care* 241 (2020), <https://doi.org/10.1186/S13054-020-03118-8>, 24 (2020) 1–6.
- [19] A. Viveiros, J. Rasmuson, J. Vu, S.L. Mulvagh, C.Y.Y. Yip, C.M. Norris, G.Y. Oudit, Sex differences in COVID-19: candidate pathways, genetics of ACE2, and sex hormones, *Am. J. Physiol. Heart Circ. Physiol.* 320 (2021), <https://doi.org/10.1152/AJPHEART.00755.2020/ASSET/IMAGES/LARGE/AJ-AHRT200091F004.JPEG>, H296–H304.
- [20] A. Freytag, J. Hoogerwerf, N. Janssen, M. Blaauw, R.-J. Hassing Rijnstate, Hospital marjan van Apeldoorn jeroen bosch hospital angele kerckhoffs jeroen bosch hospital karin verman st maartensskliniek josephine van de Maat, S. Oertelt-prigione, sex differences in the mortality of hospitalized patients with COVID-19 and non-ICU policies in The Netherlands, research gate, PREPRINT, Version 1, <https://doi.org/10.21203/rs.3.rs-1325353/v1>, 2022.
- [21] Z.G. Dessie, T. Zewotir, Mortality-related risk factors of COVID-19: a systematic review and meta-analysis of 42 studies and 423,117 patients, *BMC Infect. Dis.* 21 (2021) 1–28, <https://doi.org/10.1186/S12879-021-06536-3/FIGURES/10>.
- [22] G. Chen, D. Wu, W. Guo, Y. Cao, D. Huang, H. Wang, T. Wang, X. Zhang, H. Chen, H. Yu, X. Zhang, M. Zhang, S. Wu, J. Song, T. Chen, M. Han, S. Li, X. Luo, J. Zhao, Q. Ning, Clinical and immunological features of severe and moderate coronavirus disease 2019, *J. Clin. Invest.* 130 (2020) 2620–2629, <https://doi.org/10.1172/JCI137244>.