







## Discovery and systematic assessment of early biomarkers that predict progression to severe COVID-19 disease

Katrin Hufnagel <sup>1,8</sup>, Anahita Fathi <sup>2,3,4,5,8</sup>, Nadine Strohm <sup>1</sup>, Marco Klein<sup>1</sup>, Florian Skwirblies<sup>1</sup>, Ramy Girgis <sup>1</sup>, Christine Dahlke<sup>2,3,4</sup>, Jörg D. Hoheisel <sup>6</sup>, Camille Lowy<sup>1</sup>, Ronny Schmidt<sup>1</sup>, Anne Griesbeck<sup>1</sup>, Uta Merle<sup>7</sup>, Marylyn M. Addo <sup>2,3,4</sup> & Christoph Schröder <sup>1</sup>✉

### Abstract

**Background** The clinical course of COVID-19 patients ranges from asymptomatic infection, via mild and moderate illness, to severe disease and even fatal outcome. Biomarkers which enable an early prediction of the severity of COVID-19 progression, would be enormously beneficial to guide patient care and early intervention prior to hospitalization.

**Methods** Here we describe the identification of plasma protein biomarkers using an antibody microarray-based approach in order to predict a severe cause of a COVID-19 disease already in an early phase of SARS-CoV-2 infection. To this end, plasma samples from two independent cohorts were analyzed by antibody microarrays targeting up to 998 different proteins.

**Results** In total, we identified 11 promising protein biomarker candidates to predict disease severity during an early phase of COVID-19 infection coherently in both analyzed cohorts. A set of four (S100A8/A9, TSP1, FINC, IFNL1), and two sets of three proteins (S100A8/A9, TSP1, ERBB2 and S100A8/A9, TSP1, IFNL1) were selected using machine learning as multi-marker panels with sufficient accuracy for the implementation in a prognostic test.

**Conclusions** Using these biomarkers, patients at high risk of developing a severe or critical disease may be selected for treatment with specialized therapeutic options such as neutralizing antibodies or antivirals. Early therapy through early stratification may not only have a positive impact on the outcome of individual COVID-19 patients but could additionally prevent hospitals from being overwhelmed in potential future pandemic situations.

### Plain language summary

We aimed to identify components of the blood present during the early phase of SARS-CoV-2 infection that distinguish people who are likely to develop severe symptoms of COVID-19. Blood from people who later developed a mild or moderate course of disease were compared to blood from people who later had a severe or critical course of disease. Here, we identified a combination of three proteins that were present in the blood of patients with COVID-19 who later developed a severe or critical disease. Identifying the presence of these proteins in patients at an early stage of infection could enable physicians to treat these patients early on to avoid progression of the disease.

<sup>1</sup>Sciomics GmbH, Neckargemünd, Baden-Württemberg, Germany. <sup>2</sup>University Medical Center Hamburg-Eppendorf, Institute for Infection Research and Vaccine Development (IIRVD), Hamburg, Germany. <sup>3</sup>Bernhard-Nocht-Institute for Tropical Medicine, Department for Clinical Immunology of Infectious Diseases, Hamburg, Germany. <sup>4</sup>German Center for Infection Research, partner site Hamburg-Lübeck-Borstel-Riems, Hamburg, Germany. <sup>5</sup>University Medical Center Hamburg-Eppendorf, First Department of Medicine, Division of Infectious Diseases, Hamburg, Germany. <sup>6</sup>Division of Functional Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Baden-Württemberg, Germany. <sup>7</sup>Department of Internal Medicine IV, University Hospital Heidelberg, Heidelberg, Germany. <sup>8</sup>These authors contributed equally: Katrin Hufnagel, Anahita Fathi. ✉email: [schroeder@sciomics.de](mailto:schroeder@sciomics.de)

The outbreak of the coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was officially declared a global pandemic on March 11, 2020 by the World Health Organization (WHO). The first infection was reported by the end of December 2019 and almost two years later, the WHO reported more than 260 million confirmed cases of COVID-19, including 5.2 million deaths<sup>1</sup>. The clinical course of the disease can range from asymptomatic infection, mild and moderate disease, to severe and even critical disease courses with a case-fatality rate of up to 30% in patients older than 85 years<sup>2,3</sup>. The underlying factors causing the remarkable clinical variation of disease course are subject to intense investigation. While it is well-established that certain predisposing conditions, above all high age and, secondly, certain comorbidities<sup>4</sup>, are associated with an elevated risk for severe COVID-19, clinical outcomes still vary widely within these groups and it is not well understood which factors lead to an adverse outcome<sup>5</sup>. The identification of sensitive and specific prognostic biomarkers to predict the clinical trajectory is, therefore, crucial to guide clinical care and early intervention to ideally avoid the necessity of critical care. Several studies have described differential blood levels of various pro- or anti-inflammatory cytokines, chemokines and other proteins in sera of severely ill COVID-19 patients and suggested their use as prognostic markers for the outcome of SARS-CoV-2 infections or even as potential therapeutic targets<sup>6–9</sup>. However, most previous studies only report the analysis of a few or a single potential biomarker(s), which often lack accuracy for a clinical application. While COVID-19 vaccines have shown to be remarkably protective against severe disease, insufficient global vaccine coverage, waning immunity, and the emergence of new variants remain challenges to prevention and underline the need for effective COVID-19 treatment strategies. Immune-escape variants such as the omicron variant, furthermore, render the majority of monoclonal antibodies currently used as treatment or post-exposure prophylaxis in a wide array of patients at risk for severe COVID-19 ineffective. Due to the shortage of effective therapies against COVID-19, it is essential to select sensitive and specific biomarkers that allow in the first days of the disease and prior to an aggravation of the disease to identify those patients who are at the highest risk of developing a severe or critical course of disease and circumvent such a progression by an early treatment.

Here, we describe a study in which plasma levels of potential biomarkers were measured on antibody microarrays at an early phase of infection in COVID-19 patients which exhibited afterwards either a mild to moderate or severe to critical course of disease. Antibody microarrays are an analysis platform using antibodies immobilized on a glass slide. Patient samples were incubated on individual arrays with 1425 antibodies directed against 998 proteins. We were able to detect significantly differential plasma levels for 11 biomarkers when comparing mild to moderate and severe to critical patients during an early phase of infection in two independent cohorts. A machine learning approach revealed a combination of four biomarkers with the highest sensitivity and specificity. Individual markers could be validated by commercial ELISA assays. As the discovery approach was already completely immune-based, the biomarker combinations have the potential to be successfully translated into a clinical diagnostic tool.

## Materials & methods

**Study design.** After obtaining written informed consent, whole blood samples from SARS-CoV-2 infected individuals, who had been diagnosed by RT-PCR using a nasopharyngeal swab, were collected at University Medical Center Hamburg-Eppendorf

(1<sup>st</sup> cohort) and the Department of Gastroenterology and Infectious Diseases of University Hospital Heidelberg (2<sup>nd</sup> cohort) at different timepoints after diagnosis. Both studies were conducted according to the ethical requirements established by the Declaration of Helsinki. The 1<sup>st</sup> study was approved by local Ethics Committee of the Hamburg Medical Association (ethic consent number PV7298), while the 2<sup>nd</sup> study was approved by the local Ethics Committee of the Medical Faculty of Heidelberg University Hospital (ethic consent number S-148/2020).

All samples from the 1<sup>st</sup> cohort were classified into three phases of infection based on the onset of symptoms: acute (up to nine days after symptom onset), intermediate (10–21 days after symptom onset), and late (more than 21 days after symptom onset). All samples analyzed from the 2<sup>nd</sup> cohort were collected during the acute phase of infection.

Disease severity was classified according to the definition of the WHO-China Joint Mission on Coronavirus Disease<sup>10</sup>, which is commonly utilized as a classification, among others by the Infectious Disease Society of America (IDSA) and the Robert Koch Institute in Germany. Briefly, these categories are defined as: mild = no sign of pneumonia; moderate = radiologic evidence of pneumonia, blood oxygen saturation >93%, no supplemental oxygen therapy; severe = supplemental oxygen therapy necessary; critical = respiratory or multi-organ failure. Patient “a” of the 1<sup>st</sup> cohort represented a borderline case, who presented a poor condition with high fever, pneumonia and required hospitalization, and whose disease course was judged as clinically severe as described earlier<sup>11</sup>.

**Microarray analysis.** Sample labelling and incubation were performed as previously described in detail<sup>12–14</sup>. In brief, plasma samples from COVID-19 patients were labeled at an adjusted protein concentration for two hours with the fluorescent dye scioDye 2 (Sciomics, Neckargemünd, Germany). As reference sample, a pool composed of all samples included in each sample cohort was used and labeled with a second dye (scioDye 1). After two hours, the labeling reaction was stopped, and the buffer exchanged to Phosphate-Buffered Saline (PBS). For improved assay robustness and differentiation power, each sample was competitively incubated together with a common reference sample on one microarray slide in a reference-based dual-color approach as described in detail before<sup>15</sup>.

The 53 samples of the 1<sup>st</sup> cohort were analyzed on 53 scioCD antibody microarrays (Sciomics) targeting 351 different proteins by 517 antibodies (Supplementary Data 1). The 94 samples of the 2<sup>nd</sup> cohort were analyzed on antibody microarrays (Sciomics) targeting 998 different proteins by 1425 antibodies (Supplementary Data 2), each in four replicates. All proteins that were analyzed in the 1<sup>st</sup> cohort were also analyzed in the 2<sup>nd</sup> cohort. Array surfaces were blocked with scioBlock (Sciomics) on a Hybridization Station 4800 PRO (Tecan, Grödig, Austria) and the samples were subsequently incubated competitively with the reference sample using a dual-color approach. After incubation for three hours, the slides were thoroughly washed with 1x PBSTT (Phosphate-buffered saline containing Tween and Triton), rinsed with 0.1x PBS as well as with water and subsequently dried with nitrogen.

**Data acquisition and analysis.** Slide scanning was conducted using a Powerscanner (Tecan, GmbH, Grödig, Austria) with constant instrument laser power and photomultiplier settings. Spot segmentation was performed with GenePix Pro 6.0 (Molecular Devices, Union City, USA). The acquired raw data were analyzed using the linear models for microarray data (limma 3.42.2) package of R-Bioconductor after uploading the median

signal intensities. For normalization, a specialized invariant Lowess method was applied<sup>16</sup>. For both cohorts, a multi-factorial linear model was fitted using *limma*, integrating patient age, sex and comorbidities next to the main factor. For the first cohort, this main factor combined disease severity and phase, while for the 2<sup>nd</sup> cohort, the main factor consisted only of disease severity as the cohort exclusively consisted of acute phase samples. The factors age and comorbidities were defined in a binary manner, detailing age equal to or higher than 60 and presence of at least one comorbidity, respectively. Differences between sample groups were calculated based on the fitted group means generated by the linear model and are presented as log-fold changes (logFC) calculated for the basis 2.

**Statistics and reproducibility.** *p* values were obtained from moderated two-sided *t*-tests based on empirical Bayes moderation. All presented *p* values were adjusted for multiple testing by controlling the false discovery rate according to Benjamini and Hochberg. Proteins were defined as differentially abundant for  $|\logFC| > 0.5$  and an adjusted *p* value  $< 0.05$ . Descriptive data is depicted by range, mean and standard deviation (SD).

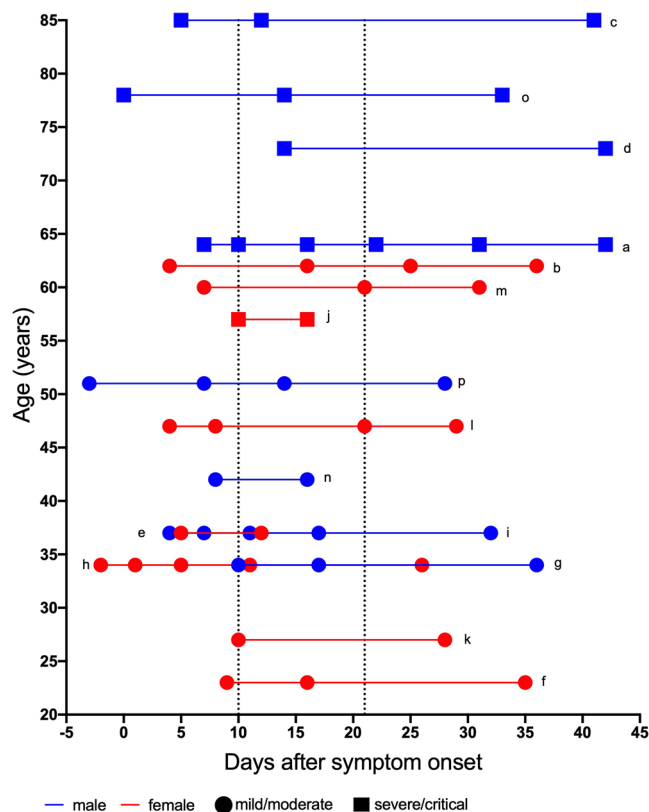
The 1<sup>st</sup> cohort comprised a total of 53 samples, including 17 acute phase samples (14 mild/moderate, 3 severe/critical), 20 intermediate phase samples (13 mild/moderate, 7 severe/critical) and 16 late phase samples (10 mild/moderate, 6 severe/critical), while the 2<sup>nd</sup> cohort consisted of a total of 94 samples, including 47 matched pairs of mild/moderate and severe/critical samples.

Signal intensity measurements for each antibody and sample were performed with four technical replicates spaced out across each microarray. The average estimated inter-replicate correlation was included into the statistical analysis.

**Machine learning.** Potential biomarker combinations were determined via linear support vector machine (linSVM) models using the SVC implementation of scikit-learn package (0.24.2) in python (3.8.10)<sup>17</sup>.

Before the machine learning process, a pre-selection of single markers was performed based on two criteria. Firstly, the 11 markers which displayed discriminative power based on logFC and *p* values obtained from the linear models of both cohorts were included. Secondly, the top 10 markers according to a preliminary linSVM coefficient criterion as detailed in Supplementary Table 1 were included for the pre-selection, resulting in a total of 21 different antibodies detecting 20 protein biomarker candidates.

From this marker set, all possible combinations of two, three and four biomarkers were tested to differentiate between patients with a critical/severe or mild/moderate course of the disease. Due to the limited size of the data set for machine learning applications, the performance of a linSVM model for a marker combination was gauged using ROC AUC estimation in a Leave-One-Out (LOO) format<sup>18</sup>. For the calculation of ROC curves, instead of binary predictions, class probabilities for the class “critical-severe”, as generated from the *predict\_proba* method of the SVC implementation, were predicted for each LOO split. This resulted in one set of predictions per biomarker combination, which was subsequently compared with the samples’ real labels to calculate the ROC AUC. In order to keep the trained models similar across the LOO folds, models were trained with a static cost parameter  $C = 1$ . The robustness of the linSVM models across the LOO folds was assessed by the coefficient of variation (CV) of the linear coefficients fitted for each fold. For the combination S100A8/A9 and CRP the CV was 12.6%, for all other individual markers and combination the CVs were in the range of 2–6% indicating robust models within the LOO process.



**Fig. 1 Patient and sample characteristics of the 1<sup>st</sup> cohort.** Characteristics (age, sex) of individuals included in this study and timepoints of sample collection after symptom onset. Letters a–p indicate the individual patients. Male patients are depicted in blue, female patients in red. Circles represent samples of mild or moderate (MM) patients, squares those of critical or severe (CS) patients. Acute phase: 14 MM and 3 CS samples; intermediate phase: 13 MM and 7 CS samples; late phase: 10 and 6 samples. Dotted vertical lines indicate 10 days and 21 days after onset of symptoms.

**Validation of biomarkers.** Concentrations of S100A8/A9 in human plasma samples were measured using the Human S100A8/S100A9 Heterodimer DuoSet Enzyme-linked Immunosorbent Assay (ELISA) (R&D Systems) according to the manufacturer’s protocol. Plasma samples were diluted 1:1000 in PBS containing 1% BSA in order to allow a measurement within the detection limits of the ELISA.

CRP was analyzed in plasma samples at the accredited central laboratory of the Heidelberg university hospital on a Siemens ADVIA Chemistry XPT System (CRP reagent kit 00829585) according to the manufacturer’s instructions.

## Results

**Study population 1<sup>st</sup> cohort.** We analyzed 53 plasma samples collected longitudinally from 16 COVID-19 patients. The cohort included eight men and women, respectively, aged between 23 and 85 years (mean: 47 years, SD: 18 years). Since it has been observed that the course of COVID-19 can deteriorate several days after the onset of first symptoms, we decided to divide the disease course into three periods based on days since the onset of first symptoms: an acute (<10 days) an intermediate (between 10 and 21 days) and a convalescent/late stage (>21 days), for which we included 17, 20 and 16 samples, respectively (see Fig. 1 for patient/sample characteristics). In all individuals, the day of first symptom onset was between February 25<sup>th</sup> and April 30<sup>th</sup>, 2020, and the cohort, therefore, mirrors an early time of the pandemic, when specific therapies against COVID-19 had not been

established. Immuno-modulating or -suppressive conditions and therapies were recorded for all participants to be able to assess whether these conditions/treatments influenced our findings. Of the 16 individuals, 2 were pregnant (patients e and h), one individual received 7.5 mg prednisolone (patient g), another one 6 mg dexamethasone daily due to underlying comorbidities (patient o) and one had received tocilizumab as a compassionate use treatment of COVID-19 (j). The remaining 11/16 patients did not have any immuno-modulating or -suppressive condition and did not receive any immunosuppressive therapy.

**Discovery of disease severity markers within the 1<sup>st</sup> cohort.** In order to identify differentially abundant proteins, human plasma samples from COVID-19 patients with either a mild or moderate (MM) or a critical or severe (CS) disease course from different disease phases (acute, intermediate, late) were analyzed on antibody microarrays targeting 351 different proteins via 517 antibodies.

100 proteins showed a significant differential abundance in CS compared to MM COVID-19 patients, from which 74 were higher and 26 lower abundant, while 417 antibodies did not show a significant difference. When classifying the samples additionally into different phases, we recorded 58 differentially abundant proteins in the acute phase (Fig. 2a), 62 in the intermediate (Fig. 2b) and 65 in the late phase (Fig. 2c). There is a broad overlap between differentially abundant proteins in CS compared to MM patients in all three phases of infections, especially between acute and intermediate phase (Fig. 2d). However, there are proteins that only show differential abundance during a specific phase of infection.

The 100 differentially abundant proteins were identified based on fitted group means generated by a linear model and thereby adjusting means for age, sex and comorbidities. This allows the identification of biomarkers associated specifically with severity and phase of disease and minimizes the chance of identifying biomarkers associated with any of these confounders.

All analyzed proteins with their respective logFC and *p* values in CS and MM patients in the three disease phases are listed in Supplementary Data 1. For each protein, the Uniprot Entry-Name and ID is listed together with the logFC and adjusted *p* values as illustrated in Fig. 2. Out of these 100 proteins, 14 top candidates were identified discriminating acute CS and acute MM patients at a  $|\logFC| > 1.0$  and a significance level of  $\text{adj.}p < 0.05$  (Fig. 3). Additionally, CD4 exhibited a high significance in intermediate phase CS patients compared to intermediate phase MM patients (Fig. 2b). Most of the selected top candidates were able to discriminate between patients with a CS and MM course of disease in the acute phase within this sample set (Fig. 3). Differences in protein levels observed in the acute phase vanished for a high number of targets within the course of disease, especially for proteins higher abundant in CS (Fig. 3). A certain heterogeneity was observed in the patient cohort in this analysis and is illustrated by the following examples. The stripchart for CD4 illustrates that this difference is mainly driven by two samples which both belong to the same patient. Similarly, differences in ICAM1 abundance might be strongly influenced by a single CS sample with extremely high ICAM1 levels, while differences in TNFR8 levels can be explained by low levels in one CS sample and additionally high levels in all samples from one MM patient in all three phases of the infection. These examples highlight that some of the differences observed within the comparatively small 1<sup>st</sup> cohort might be driven by certain patients and therefore only be present in individuals or a subpopulation. To validate the identified targets and select biomarkers, that are not only present in individuals or subpopulations, a second larger and more heterogeneous cohort was analyzed.

**Table 1 Sample characteristics of the 2<sup>nd</sup> cohort.**

	critical/severe	mild/moderate
Number of samples	47	47
Mean age (SD)	60.2 (13.4)	60.2 (12.7)
Age range	23–80	30–82
male/female	66%/34%	66%/34%
mean days after onset of symptoms (SD)	6.3 (2.1)	6.0 (2.3)

**Study population 2<sup>nd</sup> cohort – predicting a severe COVID-19 disease.** In order to assess specifically the potential to predict a severe COVID-19 disease already at an early stage of the disease, a second cohort was included in the study. The cohort consisted of 94 plasma samples from COVID-19 patients during the acute phase of disease (<10 days after the onset of first symptoms). At this stage of the disease the patients had no severe symptoms and were not in need for intensive care or hospitalisations. From these 94 patients, 47 patients later had a critical or severe course of disease, while 47 age and sex matched patients had a mild to moderate disease course only. The characteristics of these samples are summarized in Table 1. All patients survived the infection and did not receive COVID-19 specific medication prior to sample collection.

**Analysis of markers for prediction of a severe course of a COVID-19 disease within the 2<sup>nd</sup> cohort.** To identify additional protein markers and verify the findings from the initial cohort, the 94 plasma samples from the 2<sup>nd</sup> cohort were analyzed on antibody microarrays targeting 998 different proteins by 1425 antibodies.

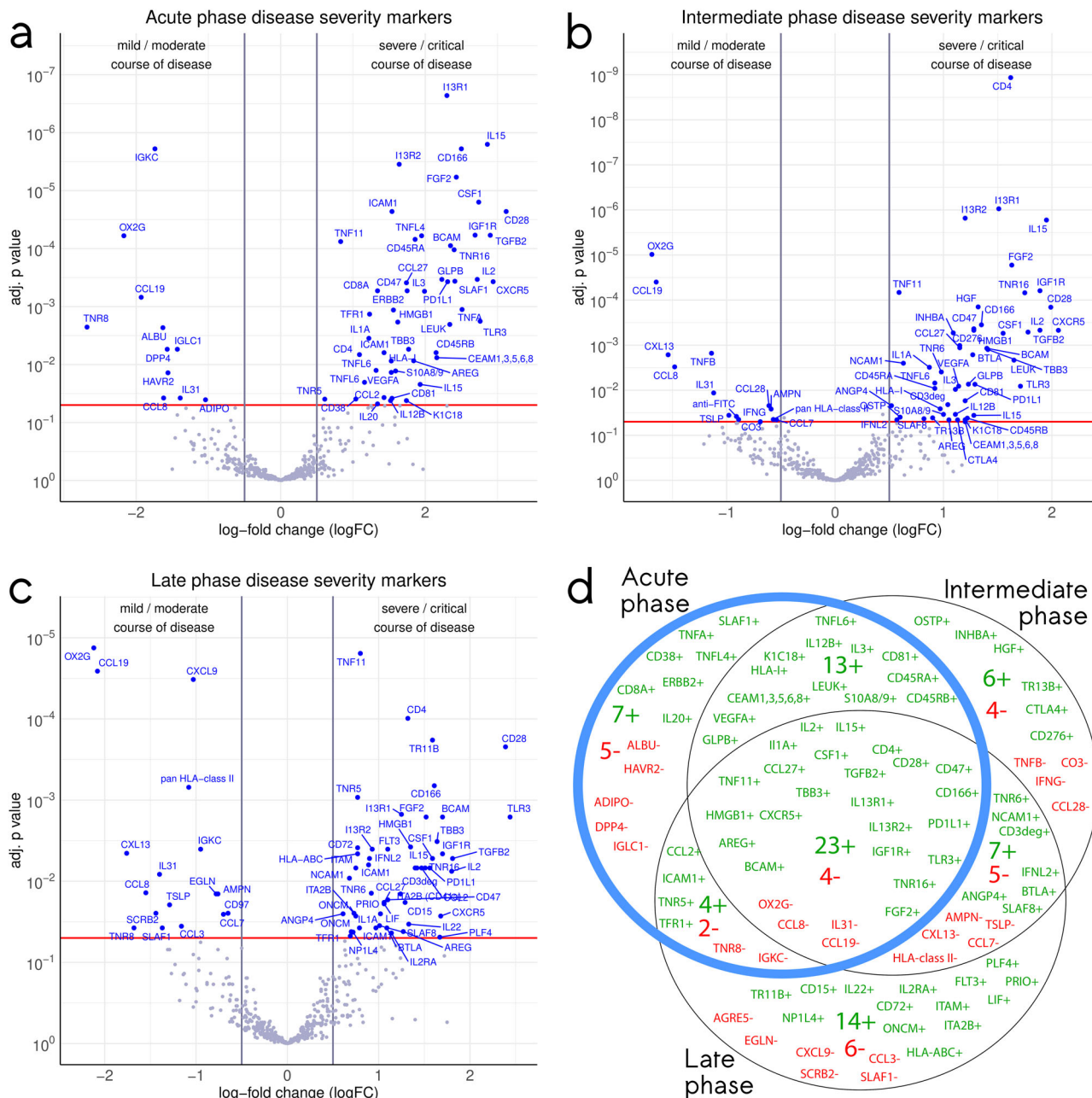
In this cohort, 51 proteins were differentially abundant in patients with a CS or MM course of the disease, from which 46 were higher abundant and five lower abundant, while 947 proteins did not show a significant difference (Fig. 4). Proteins positively associated with a severe course of the COVID-19 disease (positive logFC), such as CRP, S100A8/A9 (detected by two different antibodies), FGF2 and SLAF1, while FINC, TSP1, MMP2, IL5 and S10AD (negative logFC) were less abundant in plasma of patients with a severe or critical course of the disease. All analyzed proteins with their respective abundance in CS and MM patients are listed in Supplementary Data 2. For each protein, the Uniprot Entry-Name and ID is listed together with the logFC and adjusted *p* values.

For a mechanistic understanding of proteins altered in a CS course of the disease, the differential proteins were subjected to a protein interaction analysis using the STRING database (Supplementary Fig. 1)<sup>19</sup>. The interaction pointed specifically to a regulation of many S100 family proteins such as S100A8/A9 and S100B in combination with HMBG1. Furthermore, the Janus kinase/signal transducer and activator of transcription (JAK/STAT) pathway can be seen as one of the central axes in Supplementary Fig. 1.

**Biomarkers predictive for disease severity markers in both study cohorts.** From the 14 top candidates identified within the 1<sup>st</sup> cohort, the two proteins FGF2 and I13R2 were also differentially abundant in the 2<sup>nd</sup> cohort at a significance cutoff of  $\text{adj.}p$ -value <0.05. In total, eleven biomarkers were identified as predictive biomarkers for a severe COVID-19 disease in both study cohorts (Fig. 5, Table 2).

For the analysis of the 2<sup>nd</sup> cohort, microarrays targeting a broader range of proteins were used in order to assess additional biomarker candidates such as CRP. All biomarkers analyzed





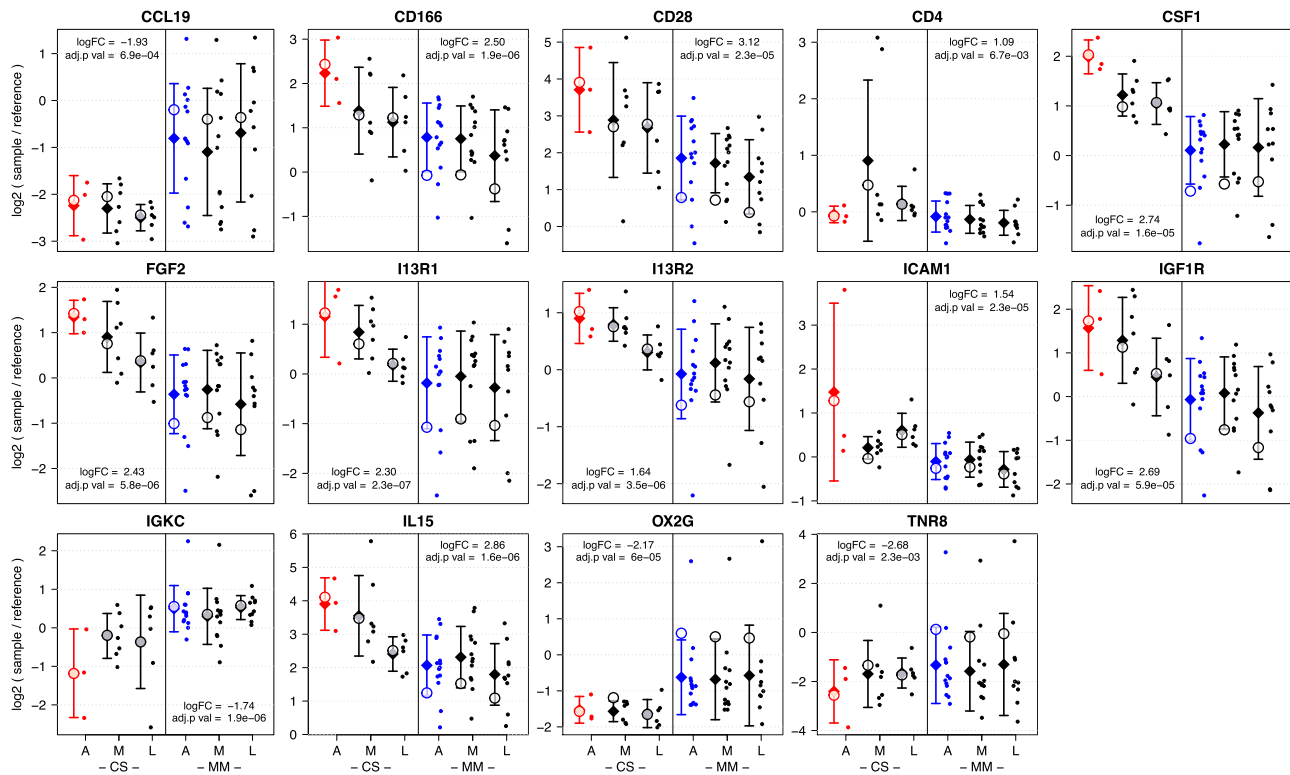
**Fig. 2** Venn diagram and volcano plots illustrating the number, degree and significance of differential protein expression in the 1<sup>st</sup> cohort. The volcano plots visualize the p values (adjusted for multiple testing) and corresponding log-fold changes (logFC) of the identified protein biomarker candidates. A significance level of adj. p value = 0.05 is indicated as a horizontal red line. Absolute logFC cutoffs of |logFC| > 0.5 are indicated as vertical lines. 53 plasma samples from 16 COVID-19 patients were analyzed and divided into an acute (a) an intermediate (b) and a late stage (c), for which we included 18, 20 and 16 samples, respectively. Proteins with a positive logFC had a higher abundance in CS samples, proteins with a negative value in MM samples. **d:** Venn diagram listing the differential proteins and their numbers in the respective phases. Green numbers and protein IDs indicate proteins more abundant in CS patients and red numbers and IDs proteins with higher abundance in MM patients.

within the 1<sup>st</sup> cohort were also included in the analysis of the 2<sup>nd</sup> cohort. Within this set, we identified the following five additional biomarkers as top candidates with a |logFC| > 1.0 and a significance level of adj. p < 0.05: CRP, FINC, TSP1, CALB1 and ANGP2. In addition, the two biomarkers MUC1 and CCL3 reached the significance threshold in the 2<sup>nd</sup> larger cohort only.

**ELISA reproduces findings of antibody array-based discovery.** To prove transferability of our findings to other assay platforms, the array data for CRP of the larger study cohort were compared

with clinical CRP data obtained at sample collection. In addition, the biomarker candidate S100A8/A9 exhibiting the second strongest discriminative power was chosen for such inter-assay comparison using a commercially available ELISA kit. Antibody array data and ELISA data exhibited a Pearson correlation coefficient of 0.905 for S100A8/A (Fig. 6c) and of 0.955 for CRP (Fig. 6d).

Also, in terms of discrimination power, the performance of the antibody array platform could be reproduced by ELISA. By applying a cutoff of 1.45 μg/ml for S100A8/A9 measured by ELISA, a specificity of 83% and a sensitivity of 89% were



**Fig. 3 Stripcharts representing individual array values for all proteins selected as top candidates in the 1<sup>st</sup> cohort.** Each protein is measured by four replicate spots per array and is represented by their mean. The y-axis illustrates the log<sub>2</sub> ratio of the individual samples and a reference sample while the x-axis is divided based on clinical course of disease of the patient (CS and MM) as well as phase of infection. 53 plasma samples from 16 COVID-19 patients were analyzed and divided into an acute (A) an intermediate (M) and a late stage (L), for which we included 18, 20 and 16 samples, respectively. Acute CS and MM samples are highlighted in red and blue respectively. Diamonds indicate arithmetic sample group means. Whiskers indicate one standard deviation, calculated based on arithmetic means. Empty circles indicate the group coefficients fitted by the linear model with additional factors sex, age and comorbidities, which were used for logFC and *p* value calculation.

achieved. A cut off of 14.5 µg/ml for CRP measured by ELISA resulted in a specificity of 73% and a sensitivity of 89%.

**Biomarker combinations selected via machine learning for accurate disease severity prediction.** In order to assess the diagnostic accuracy of the two top protein biomarker candidates, receiver operating characteristic (ROC) curves were generated for both the antibody array derived data and the ELISA data. For CRP, the area under the curve (AUC) was 0.837 based on the antibody array data and 0.866 for the ELISA data (Fig. 7). The biomarker S100A8/A9 can predict a severe or critical COVID-19 disease at an AUC of 0.827 based on the antibody array data and 0.886 based on the ELISA data (Fig. 7). At a sensitivity of 90%, a specificity of 59.6% can be reached for CRP and of 48.9% for S100A8/A9 based on the antibody array data. ROC curves for the two proteins are in good coherence for ELISA and antibody array data with the ELISA data slightly outperforming the data derived from highly multiplex antibody arrays (Fig. 7).

In order to further improve the diagnostic accuracy, especially in terms of sensitivity, potential biomarker combinations were determined via machine learning models. As linear support vector machines (linSVM) are easily interpretable and can be easily transferred into a diagnostic assay format, in our approach we focused and described models selected based on linSVMs rather than tree-based models like XGBoost.

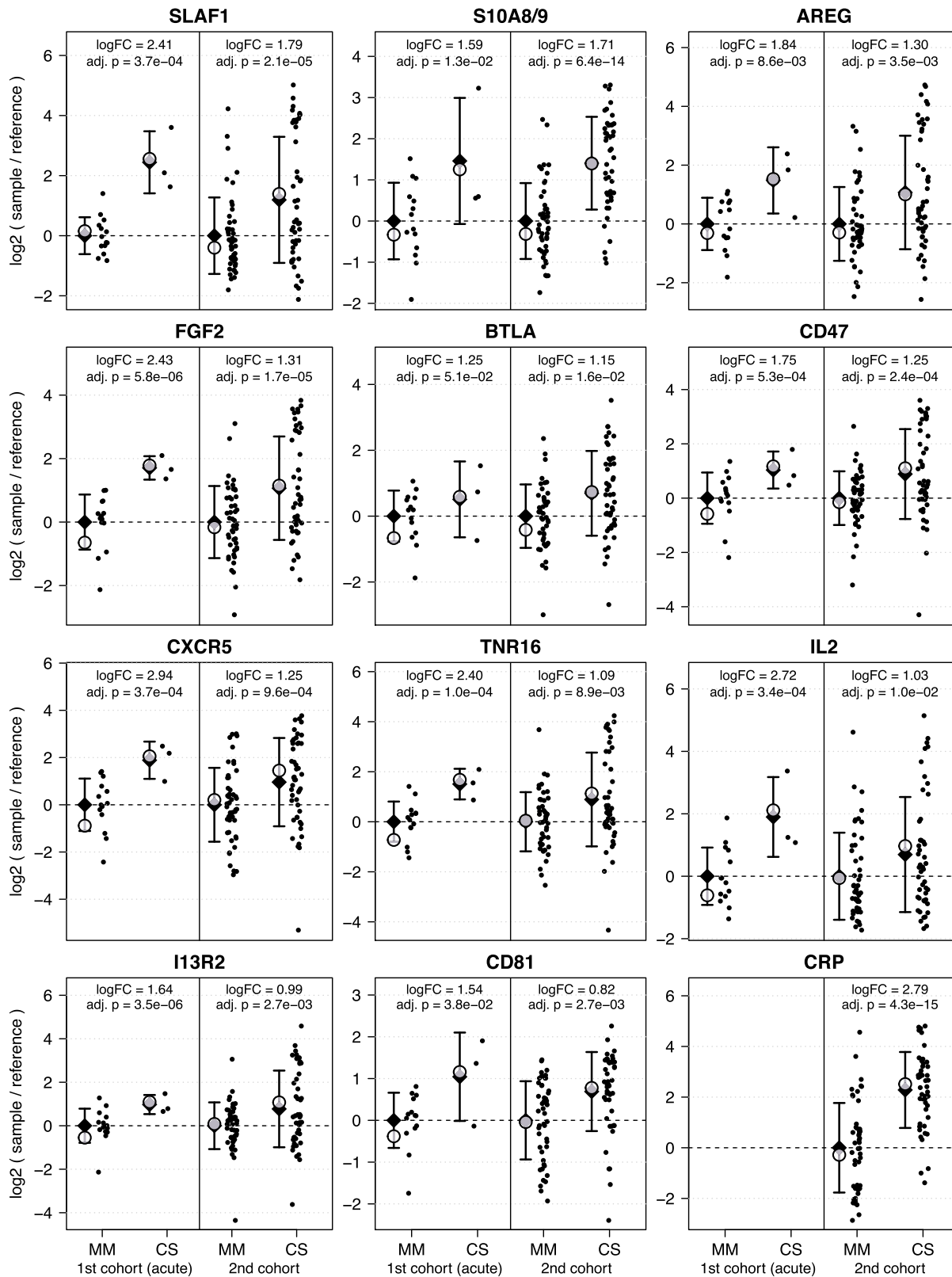
From the many evaluated linSVM signatures, we present the ten models yielding the highest performance based on their AUC for signature lengths of two, three and four proteins, respectively, within Supplementary Table 2 and highlighted four of these

models within Table 3 and Fig. 7. The identified models were compared to a combination of the top single protein biomarker candidates S100A8/A9 and CRP as well as their combination. The combination of the two biomarkers revealed an AUC of 0.830 and a specificity of 46.8% at a sensitivity of 90% and thereby a similar performance as the individual markers. For other combinations selected from the linSVM models, the overall performance and especially the sensitivity could be improved. A combination of S100A8/A9 with the protein TSP1 yielded an AUC of 0.872, thereby showing the highest AUC of all marker panels consisting of two biomarkers. If ERBB2 is added to this marker panel, the AUC is further improved to 0.898. A higher AUC of 0.913 was reached with one other panel of the three biomarkers S100A8/A9, TSP1 and IFNL1. A further addition of FINC to this marker panel increases the AUC to 0.928 with a slightly lower specificity at a sensitivity of 95% as compared to the two selected three protein biomarker panels. For a detection of a severe disease course using the four marker panels highlighted in Fig. 7, a sensitivity of 90% could be reached at specificities of 78.7%, 80.9%, 78.7% and 83.0%, respectively (Table 3), clearly outperforming all individual protein markers as well as the combination of S100A8/A9 and CRP.

## Discussion

The presented study aimed to find sensitive and specific biomarkers to identify patients at high risk of developing a severe or critical course of COVID-19 within an early phase of infection and prior to the onset of severe symptoms. Antibody microarray analysis of samples, taken during the acute phase of COVID-19





**Fig. 5 Biomarkers to predict a severe / critical disease in both study cohorts.** The y-axis illustrates log<sub>2</sub>(sample / reference) values after subtracting the group mean of the respective MM samples per cohort/protein, thus setting the mean value of MM samples as a baseline. Within the 1<sup>st</sup> cohort 18 samples (MM = 17; CS = 3) were analyzed, while 94 sample (MM = 47; CS = 47) were analyzed within the 2<sup>nd</sup> cohort. The x-axis is divided based on clinical course of disease of the patient (CS and MM). Only acute CS and MM samples are shown. Diamonds indicate arithmetic sample group means. Whiskers indicate one standard deviation, calculated based on the arithmetic means. Empty circles indicate the group coefficients fitted by the linear model with additional factors sex, age and comorbidities, which were used for logFC and p value calculation.



**Table 2 Most important identified biomarkers within at least one of both cohorts.**

	Target	Uniprot Name	Uniprot ID	1 <sup>st</sup> cohort		2 <sup>nd</sup> cohort		
				logFC	adj.p-val	logFC	adj.p-val	
common targets	S100A8/A9	S10A8_HUMAN/S10A9_HUMAN	P05109/P06702	1.59	0.013	1.71	$6.4 \times 10^{-14}$	
	FGF2	FGF2_HUMAN	P09038	2.43	$5.8 \times 10^{-6}$	1.31	$1.7 \times 10^{-5}$	
	SLAF1	SLAF1_HUMAN	Q13291	2.41	$3.7 \times 10^{-4}$	1.79	$2.1 \times 10^{-5}$	
	CD47	CD47_HUMAN	Q08722	1.75	$5.3 \times 10^{-4}$	1.25	$2.4 \times 10^{-4}$	
	CXCR5	CXCR5_HUMAN	P32302	2.94	$3.7 \times 10^{-4}$	1.25	$9.6 \times 10^{-4}$	
	I13R2	I13R2_HUMAN	Q14627	1.64	$3.5 \times 10^{-6}$	0.99	$2.7 \times 10^{-3}$	
	CD81	CD81_HUMAN	P01889	1.54	0.038	0.82	$2.7 \times 10^{-3}$	
	AREG	AREG_HUMAN	P31997	1.84	$8.6 \times 10^{-3}$	1.30	$3.5 \times 10^{-3}$	
	TNFR16	TNFR16_HUMAN	P08138	2.40	$1.0 \times 10^{-4}$	1.09	$8.9 \times 10^{-3}$	
	IL2	IL2_HUMAN	P60568	2.72	$3.4 \times 10^{-4}$	1.03	0.010	
	BTLA	BTLA_HUMAN	Q7Z6A9	1.27*	$1.6 \times 10^{-3}$ *	1.15	0.016	
	important targets 1st cohort	I13R1	I13R1_HUMAN	P78552	2.30	$2.3 \times 10^{-7}$	ns	ns
		IL15	IL15_HUMAN	P40933	2.86	$1.6 \times 10^{-6}$	ns	ns
		CD166	CD166_HUMAN	Q13740	2.50	$1.9 \times 10^{-6}$	ns	ns
IGKC		IGKC_HUMAN	P01834	-1.74	$1.9 \times 10^{-6}$	ns	ns	
CSF1		CSF_HUMAN	P09603	2.74	$1.6 \times 10^{-5}$	ns	ns	
CD28		CD28_HUMAN	P10747	3.12	$2.3 \times 10^{-5}$	ns	ns	
IGF1R		IGF1R_HUMAN	P08069	2.69	$5.9 \times 10^{-5}$	ns	ns	
OX2G		OX2G_HUMAN	P41217	-2.17	$6.0 \times 10^{-5}$	ns	ns	
CCL19		CCL19_HUMAN	Q99731	-1.93	$6.9 \times 10^{-4}$	ns	ns	
important targets 2nd cohort		CRP	CRP_HUMAN	P02741	na	na	2.79	$4.3 \times 10^{-15}$
	FINC	FINC_HUMAN	P02751	na	na	-0.98	$3.7 \times 10^{-5}$	
	TSP1	TSP1_HUMAN	P07996	na	na	-0.55	$1.7 \times 10^{-4}$	
	MUC1	MUC1_HUMAN	P15941	ns	ns	1.74	$1.0 \times 10^{-4}$	
	CALB1	CALB1_HUMAN	P05937	na	na	1.37	$8.5 \times 10^{-4}$	
	CCL3	CCL3_HUMAN	P10147	ns	ns	1.49	$2.7 \times 10^{-3}$	
	ANGP2	ANGP2_HUMAN	O15123	na	na	1.28	$6.3 \times 10^{-3}$	

\*BTLA logFC and p-values (p-val.) for the 1<sup>st</sup> cohort are from the intermediate phase of infection. ns not significant. na not available; target was not analyzed in this cohort.

This might be due to the fact that some differential protein levels have been driven by individuals. Another explanation might be that these proteins are connected to age or certain comorbidities and thereby not exhibit a significance in the 2<sup>nd</sup> cohort controlling for these potential confounding factors with a case-control design.

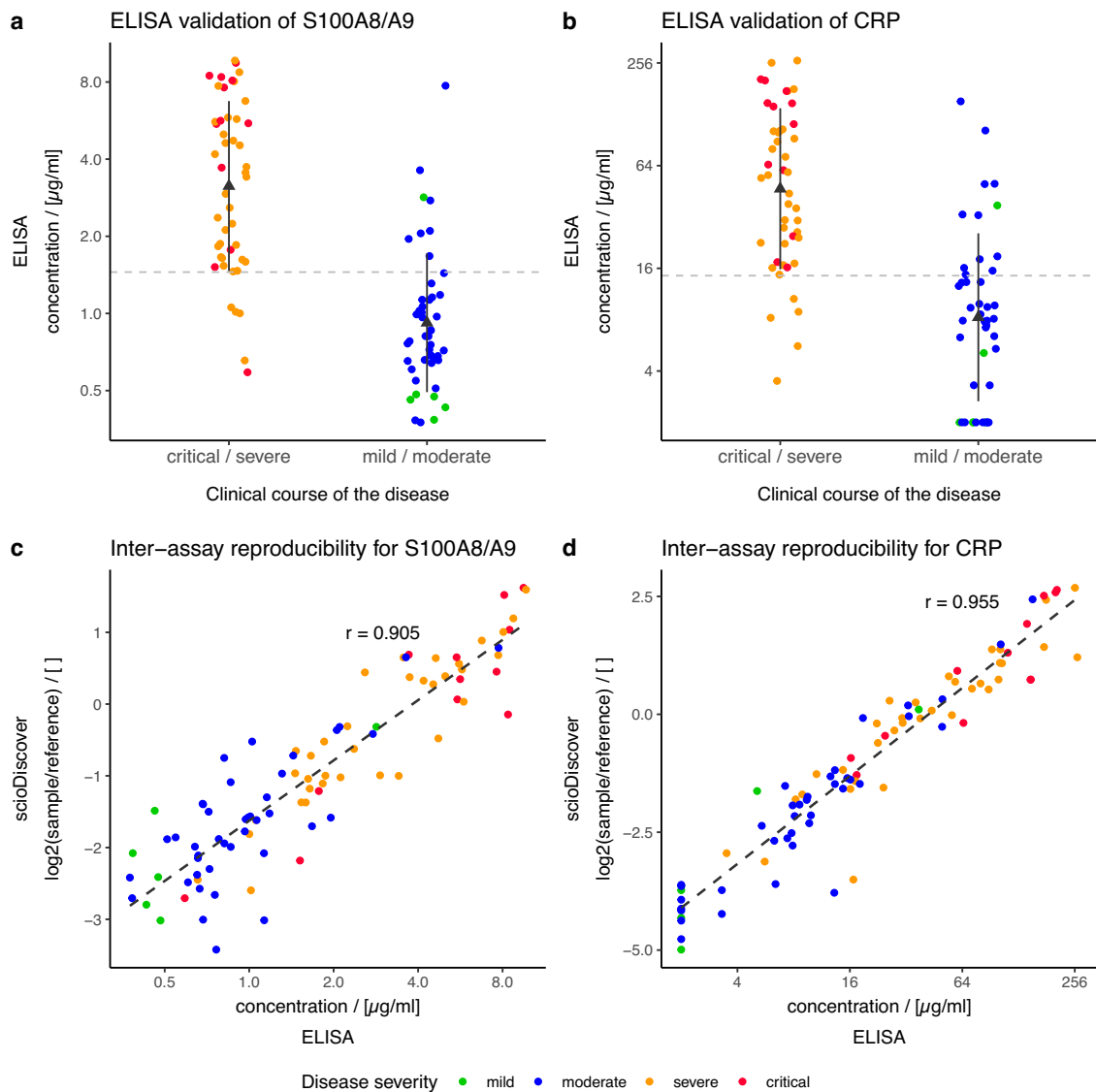
All identified proteins might be promising biomarkers to detect individuals at a high risk for developing severe COVID-19 during an early phase of infection. Up-regulated protein levels in CS patients, observed in the acute phase, showed decreasing abundance during the intermediate and late phase for some targets when analyzing longitudinal samples from the 1<sup>st</sup> cohort. Since all CS patients within this study survived COVID-19, these markers, as well as markers identified in the intermediate phase of the disease might be promising markers for therapy monitoring.

For the analysis of the 2<sup>nd</sup> cohort, antibody microarrays targeting a broader range of proteins were used in order to gain more comprehensive insights into plasma protein profiles connected with the onset of severe COVID-19. A protein interaction analysis of the biomarkers to predict a critical or severe course of the disease, revealed the JAK/STAT pathway as one of the central axes in the protein interaction network. A JAK/STAT activation is frequently seen in infectious diseases and a sign for immune cell activation and proliferation. Drugs targeting the JAK/STAT pathway are currently under consideration as treatment options in COVID-19<sup>42</sup>. In addition, the protein interaction analysis revealed a regulation of many S100 family proteins such as S100A8/A9 and S100B in combination with HMBG1. Other mechanisms and pathways identified in the protein interaction network of the biomarker candidates are part of a systemic response to a severe viral infection. This systemic response

comprises the activation of many immune cell types as well as complex regulations of growth factors and soluble inflammatory mediators, including IL15, I13R2 (HGNC: IL13RA2), CD47 and TSP1 (HGNC: THBS1).

As a response protein to inflammation, CRP was identified in the 2<sup>nd</sup> cohort as one of the top individual markers. Although CRP is a rather unspecific inflammation marker, it has widely been published as a prognostic biomarker for COVID-19 progression<sup>43</sup>, and we were able to reproduce these published data by our array platform as well as by an ELISA performed in a routine clinical diagnostics laboratory. Scotto et al. showed in correlation with our findings that MUC1 (Mucin-1) is a reliable indicator of pulmonary function, reported an association with poorer outcome and death in COVID-19 patients and therefore suggested MUC1 as a sensitive parameter to stratify the risk of severe respiratory failure and death in COVID-19 patients<sup>44</sup>. As in our study, enhanced expression of CCL3<sup>45,46</sup> and ANGP2<sup>47</sup> were also described as predictive factors in differentiating COVID-19 patients and determining severity of disease.

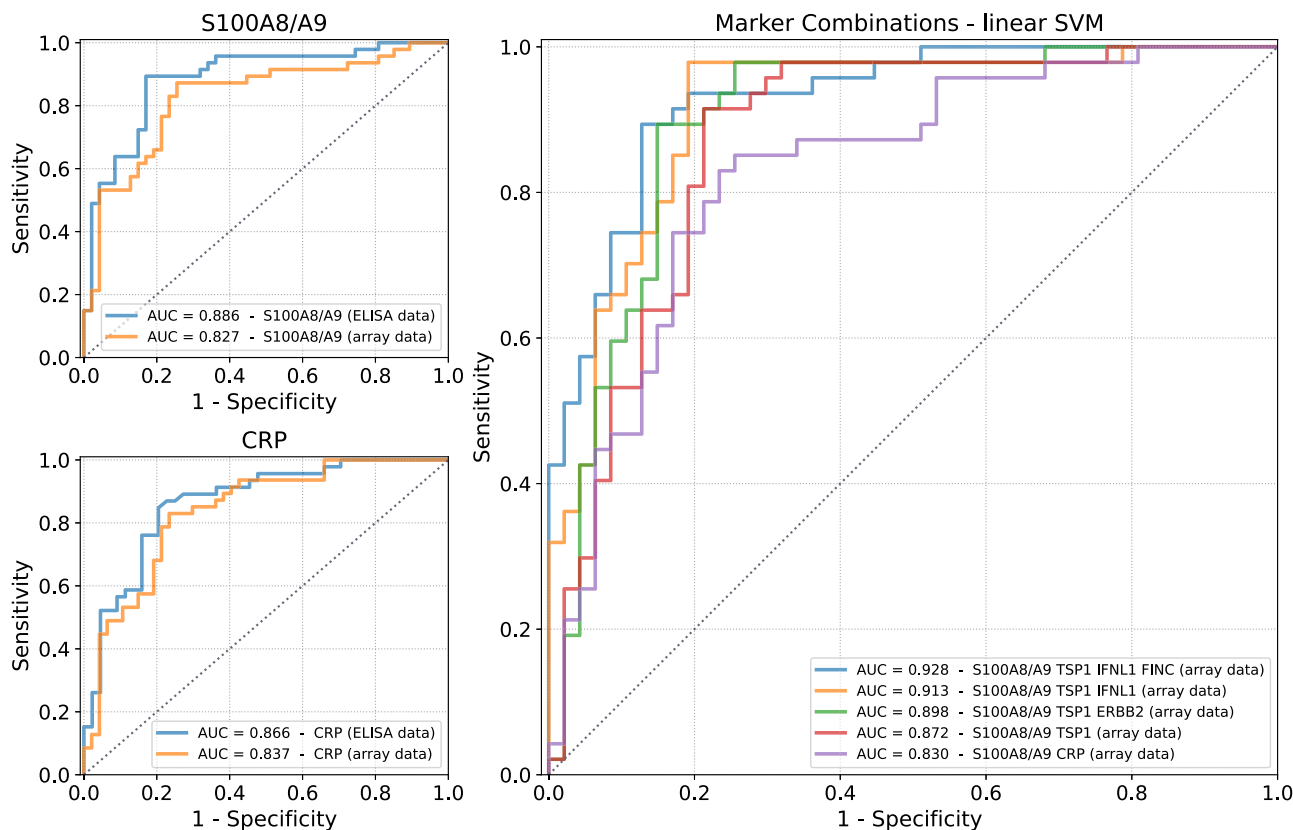
Individual biomarker candidates with the highest log-fold change in the 2<sup>nd</sup> cohort exhibited also the highest diagnostic accuracy with an AUC of 0.83. Still, this performance might not be sensitive enough, especially in a larger study population with a higher number of mild and moderate cases. Therefore, machine learning approaches were applied to select ideal marker combinations for improved AUC as well as sensitivity. To this end, linear as well as decision tree approaches were tested. While tree-based approaches based on XGBoost had a slightly better performance, we decided to focus on linear approaches based on linSVM, as linear models can be more easily reproduced for the setup of a multiplex assay aiming at in vitro diagnostic certification.



**Fig. 6 Validation of microarray data.** Plasma samples from 94 COVID-19 patients during the acute phase of disease (2<sup>nd</sup> cohort) were analyzed by ELISA, to validate differentially abundant proteins between patients with either a mild or moderate (MM,  $n = 47$ ) or a critical or severe (CS,  $n = 47$ ) disease course. Stripcharts representing individual S100A8/A9 (**a**) and CRP (**b**) ELISA measurements. The y-axis displays the  $\log_2$  of the measured protein concentration while the x-axis is divided based on the later clinical course of disease of the patient (CS and MM). Triangles and whiskers indicate means and one standard deviation of the sample groups with critical/severe or mild/moderate course of the disease respectively. Possible cut-offs with a sensitivity of 89% are indicated by dotted grey lines. **c, d** Scatter plots demonstrate a high correlation between discovery antibody microarray data (y-axis) and ELISA validation (x-axis) with Pearson's  $r > 0.9$ .

Due to the larger sample size of the 2<sup>nd</sup> cohort and the divergence in selection of samples in the two studies – matched mild/moderate and severe/critical cases in respect to age, gender, and days after onset of symptoms only in the 2<sup>nd</sup> cohort –, the 2<sup>nd</sup> cohort was used for the linSVM approach to select biomarker combinations discriminating between patients with CS and MM disease courses at a higher accuracy than individual protein biomarkers. A combination of the four proteins S100A8/9, FINC, IFNL1 and TSP1 (AUC = 0.928) as well as of the three proteins S100A8/A9, TSP1 and ERBB2 (AUC = 0.898) but also of the two protein biomarkers S100A8/A9 and TSP1 (AUC = 0.872) outperformed individual biomarkers such as S100A8/A9 (AUC: 0.827) or CRP (AUC: 0.837) clearly. FINC and TSP1 levels were not measured within the 1<sup>st</sup> cohort but were significantly higher in MM patients as compared to CS patients within the 2<sup>nd</sup> cohort. From the proteins with specific importance in the biomarker

signatures identified by machine learning, FINC (Fibronectin) has not yet been published as a biomarker for severe COVID-19, however it has been identified in inflammation and sepsis as a negative acute-phase protein, with a low level of FINC indicating a poor prognosis for a patient<sup>48</sup>. For Thrombospondin-1 (TSP1) we did not identify any report as a COVID-19 disease severity biomarker, however TSP1 protein levels were shown to be up-regulated in the serum of infected asymptomatic individuals as compared to negative individuals<sup>49</sup>. In lung cancer patients, lower TSP1 levels are associated with a higher lymph node involvement in lung cancer<sup>50</sup>. IFNL1 levels were slightly higher in MM patients as compared to CS patients within the 2<sup>nd</sup> cohort. Although differences in IFNL1 were not significant as an individual biomarker, IFNL1 was selected by the linear SVM approach selecting combinations of biomarkers to improve sensitivity and specificity as compared to a single biomarker. Lambda



**Fig. 7 Estimation of diagnostic accuracy for disease severity prediction.** For the individual protein biomarkers S100A8/A9 and CRP, ROC curves of ELISA data as well as coherent multiplex antibody array data were aligned. Marker combinations of 2–4 proteins were selected from linSVM models, which outperform the individual biomarker candidates and exhibit a high accuracy with an AUC of up to 0.928. Area under the ROC curve (AUC) is presented for each biomarker or combination.

**Table 3 Specificities at given sensitivities for selected individual biomarkers and biomarker combinations.**

Biomarker/Biomarker Combination	AUC	Specificity at a Sensitivity of			
		95%	90%	85%	80%
S100A8/A9	0.827	19.1%	48.9%	74.5%	76.6%
CRP	0.837	34.0%	59.6%	70.2%	76.6%
S100A8/A9 + CRP	0.830	46.8%	46.8%	74.5%	76.6%
S100A8/A9 + TSP1	0.872	70.2%	78.7%	78.7%	80.9%
S100A8/A9 + TSP1 + IFNL1	0.913	80.9%	80.9%	83.0%	83.0%
S100A8/A9 + TSP1 + ERBB2	0.898	74.5%	78.7%	85.1%	85.1%
S100A8/A9 + TSP1 + FINC + IFNL1	0.928	63.8%	83.0%	87.2%	87.2%

IFNs such as IFNL1 are produced as a response to a viral infection and act in mucosal barriers, e.g. in the respiratory tract. The concentrations of lambda IFNs induced in the serum of non-ICU COVID-19 patients were shown to be higher than in ICU COVID-19 patients<sup>51</sup>. Additionally, it was shown that IFNL1 restricts SARS-CoV-2 replication in cellular models of viral infection when administered preventively and/or therapeutically<sup>52</sup>. In MERS-CoV, lambda IFN greatly limits viral replication, and may be a key cytokine for better therapeutic outcomes against MERS-CoV infection in the respiratory tract<sup>53</sup>.

All machine learning approaches were performed using data based on log-ratios of sample and reference as derived from the antibody array platform. Such multiplex antibody array data

proved very valuable for machine learning, as, contrary to many other platforms, there is no issue of missing data.

As demonstrated for the two individual top biomarker candidates S100A8/A9 and CRP, the array data could be very well correlated ( $r > 0.9$ ) and reproduced by commercial ELISA. In a direct comparison by a side-to-side ROC analysis of array and ELISA data, the ROC curves for S100A8/A9 and CRP originating from the ELISA data slightly outperformed the antibody array data. These differences are likely a result of a combination of factors, such as a more complex experimental layout (reference-based), different laboratory procedures, in particular adjustment of bulk protein concentrations for plasma samples, and data normalization on the side of the array data.

Nonetheless, these findings indicate a very good transferability of the results of this study to routine platforms for protein analysis with a low risk of false-positive findings. This direct transferability of results to other immune based platforms used in routine diagnostics such as ELISA is a major advantage of antibody array-based proteomics. Additional advantages of the antibody microarray platform include the highly simultaneous screening of several hundred (1<sup>st</sup> cohort) and more than a thousand (2<sup>nd</sup> cohort) proteins from less than fifty microliters of plasma sample in a comparably high throughput, and sensitivity levels enabling a robust cytokine profiling from complex samples without the need of an intensive sample depletion and fractionation as required in mass spectrometry-based approaches.

Based on the Sciomics antibody microarray platform we were able to identify plasma protein biomarkers to predict a severe COVID-19 disease, selected protein biomarker combinations with an accuracy suitable for a clinical application and validated

two of the individual protein biomarkers by commercial assays. The identified proteins are promising biomarkers to predict disease severity during an early phase of SARS-CoV-2 infection. The evidence that is available to date suggests that these markers are connected to mechanisms involved in disease progression to severe COVID-19 as well as in other infectious disease models. Individually or collectively, these markers can identify patients at high risk of developing a severe and critical course of disease and can inform treatment and clinical care choices.

**Reporting summary.** Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The authors declare that all logFC data, generated during the microarray analysis are available within the paper and its supplementary information files (Supplementary Data 1 and Supplementary Data 2). Raw red and green signal intensities and the resulting normalised M-values of both cohorts are deposited at ArrayExpress (1st cohort: E-MTAB-12779 Discovery and systematic assessment of early biomarkers that predict progression to severe COVID-19 disease - 1st cohort; 2nd cohort: E-MTAB-12777 Discovery and systematic assessment of early biomarkers that predict progression to severe COVID-19 disease - 2nd cohort). Data for reproducing the figures is available within the supplementary data 3.

### Code availability

Microarray raw image files were processed using GenePix Pro 6.0 and statistical analysis was performed using the limma package (3.42.2) in R (3.6.3) as outlined in detail in a publication describing the data analysis routines for antibody microarray analyses<sup>16</sup>. Machine learning procedures were performed using the scikit-learn package (0.24.2) in python (3.8.10) according to the methodology described in the Materials & Methods section. The code was published and can be accessed via the following DOI: 10.5281/zenodo.7759895<sup>18</sup>.

Received: 15 February 2022; Accepted: 31 March 2023;  
Published online: 12 April 2023

### References

1. Coronavirus disease (COVID-19) – World Health Organization. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
2. CDC COVID-19 Response Team. Severe outcomes among patients with coronavirus disease 2019 (COVID-19) - United States, February 12-March 16, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 343–346 (2020).
3. Schilling, J. et al. Krankheitsschwere der ersten COVID-19-Welle in Deutschland basierend auf den Meldungen gemäß Infektionsschutzgesetz. (2020) <https://doi.org/10.25646/7169>.
4. Ständige Impfkommission: Empfehlungen der Ständigen Impfkommission (STIKO) beim Robert Koch-Institut. *Epid Bull* **4**, 3–67 <https://doi.org/10.25646/9285.3> (2022).
5. Marin, B. G. et al. Predictors of COVID-19 severity: a literature review. *Rev. Med. Virol.* **31**, 1–10 (2021).
6. Bénard, A. et al. Interleukin-3 is a predictive marker for severity and outcome during SARS-CoV-2 infections. *Nat. Commun.* **12**, 1112 (2021).
7. Shen, B. et al. Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* **182**, 59–72.e15 (2020).
8. Blot, M. et al. The dysregulated innate immune response in severe COVID-19 pneumonia that could drive poorer outcome. *J. Transl. Med.* **18**, 457 (2020).
9. Abers, M. S. et al. An immune-based biomarker signature is associated with mortality in COVID-19 patients. *JCI Insight* **6**, e144455 (2021).
10. WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). <https://www.who.int/docs/default-source/coronavirus/who-china-joint-mission-on-covid-19-final-report.pdf> (2020).
11. Heidepriem, J. et al. Longitudinal development of antibody responses in covid-19 patients of different severity with elisa, peptide, and glycan arrays: an immunological case series. *Pathog. Basel Switz.* **10**, 438 (2021).
12. Janša, V. et al. Proteomic analysis of peritoneal fluid identified COMP and TGFBI as new candidate biomarkers for endometriosis. *Sci. Rep.* **11**, 20870 (2021).
13. Nijaguna, M. B. et al. Definition of a serum marker panel for glioblastoma discrimination and identification of Interleukin 1 $\beta$  in the microglial secretome as a novel mediator of endothelial cell survival induced by C-reactive protein. *J. Proteom.* **128**, 251–261 (2015).
14. Schröder, C. et al. Plasma protein analysis of patients with different B-cell lymphomas using high-content antibody microarrays. *PROTEOMICS – Clin. Appl.* **7**, 802–812 (2013).
15. Schröder, C. et al. Dual-color proteomic profiling of complex samples with a microarray of 810 cancer-related antibodies. *Mol. Cell. Proteom. MCP* **9**, 1271–1280 (2010).
16. Sill, M., Schröder, C., Hoheisel, J. D., Benner, A. & Zucknick, M. Assessment and optimisation of normalisation methods for dual-colour antibody microarrays. *BMC Bioinform.* **11**, 556 (2010).
17. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
18. Hufnagel, Katrin et al. *Machine Learning Code for publication 'Discovery and systematic assessment of early biomarkers that predict progression to severe COVID-19 disease'*. (Zenodo, 2023). <https://doi.org/10.5281/ZENODO.7759895>.
19. STRING: functional protein association networks. <https://string-db.org/>.
20. Khinda, J. et al. Association between markers of immune response at hospital admission and COVID-19 disease severity and mortality: a meta-analysis and meta-regression. *J. Med. Virol.* **93**, 1078–1098 (2021).
21. Elshazli, R. M. et al. Diagnostic and prognostic value of hematological and immunological markers in COVID-19 infection: a meta-analysis of 6320 patients. *PLoS One* **15**, e0238160 (2020).
22. Mahler, M., Meroni, P.-L., Infantino, M., Buhler, K. A. & Fritzler, M. J. Circulating calprotectin as a biomarker of COVID-19 severity. *Expert Rev. Clin. Immunol.* **17**, 431–443 (2021).
23. Chen, L. et al. Elevated serum levels of S100A8/A9 and HMGB1 at hospital admission are correlated with inferior clinical outcomes in COVID-19 patients. *Cell. Mol. Immunol.* 1–3 (2020) <https://doi.org/10.1038/s41423-020-0492-x>.
24. Shi, H. et al. Neutrophil calprotectin identifies severe pulmonary disease in COVID-19. *J. Leukoc. Biol.* **109**, 67–72 (2021).
25. Silvina, A. et al. Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell* **182**, 1401–1418.e18 (2020).
26. Kaya, T. et al. Serum calprotectin as a novel biomarker for severity of COVID-19 disease. *Ir. J. Med. Sci.* 1–6 (2021) <https://doi.org/10.1007/s11845-021-02565-8>.
27. Petrey, A. C. et al. Cytokine release syndrome in COVID-19: innate immune, vascular, and platelet pathogenic factors differ in severity of disease and sex. *J. Leukoc. Biol.* **109**, 55–66 (2021).
28. Meini, S., Giani, T. & Tascini, C. Intussusceptive angiogenesis in Covid-19: hypothesis on the significance and focus on the possible role of FGF2. *Mol. Biol. Rep.* **47**, 8301–8304 (2020).
29. Zhang, Y. et al. Identification of monocytes associated with severe COVID-19 in the PBMCs of severely infected patients through single-cell transcriptome sequencing. *Eng. Beijing China* (2021) <https://doi.org/10.1016/j.eng.2021.05.009>.
30. Shenoy, S. SARS-CoV-2 (COVID-19), viral load and clinical outcomes; lessons learned one year into the pandemic: a systematic review. *World J. Crit. Care Med.* **10**, 132–150 (2021).
31. Consortium, Co.-19 M. B. At. (COMBAT) et al. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *medRxiv* 2021.05.11.21256877 (2021) <https://doi.org/10.1101/2021.05.11.21256877>.
32. Chen, C., Shi, L., Li, Y., Wang, X. & Yang, S. Disease-specific dynamic biomarkers selected by integrating inflammatory mediators with clinical informatics in ARDS patients with severe pneumonia. *Cell Biol. Toxicol.* **32**, 169–184 (2016).
33. Winn, B. J. Is there a role for insulin-like growth factor inhibition in the treatment of COVID-19-related adult respiratory distress syndrome? *Med. Hypotheses* **144**, 110167 (2020).
34. Vaine, C. A. & Soberman, R. J. The CD200–CD200R1 inhibitory signaling pathway. *Adv. Immunol.* **121**, 191–211 (2014).
35. Ceribelli, A. et al. Recommendations for coronavirus infection in rheumatic diseases treated with biologic therapy. *J. Autoimmun.* **109**, 102442 (2020).
36. Chitu, V. & Stanley, E. R. Colony-stimulating factor-1 in immunity and inflammation. *Curr. Opin. Immunol.* **18**, 39–48 (2006).
37. Quartuccio, L. et al. Interleukin 6, soluble interleukin 2 receptor alpha (CD25), monocyte colony-stimulating factor, and hepatocyte growth factor linked with systemic hyperinflammation, innate immunity hyperactivation, and organ damage in COVID-19 pneumonia. *Cytokine* **140**, 155438 (2021).
38. Masedunskas, A. et al. Activated leukocyte cell adhesion molecule is a component of the endothelial junction involved in transendothelial monocyte migration. *FEBS Lett* **580**, 2637–2645 (2006).
39. Mukherjee, M. et al. CD6-ALCAM Pathway is elevated in patients with severe asthma. *Eur. Respir. J.* **56**, 5283 (2020).



40. Rébillard, R.-M. et al. Identification of SARS-CoV-2-specific immune alterations in acutely ill patients. *J. Clin. Invest.* **131**, e145853 <https://doi.org/10.1172/JCI145853> (2021).
41. Wang, F. et al. The laboratory tests and host immunity of COVID-19 patients with different severity of illness. *JCI Insight* **5**, e137799 (2020).
42. Satarker, S. et al. JAK-STAT Pathway inhibition and their implications in COVID-19 therapy. *Postgrad. Med.* **133**, 489–507 (2021).
43. Stringer, D. et al. The role of C-reactive protein as a prognostic marker in COVID-19. *Int. J. Epidemiol.* **50**, 420–429 (2021).
44. Scotto, R. et al. Serum KL-6 could represent a reliable indicator of unfavourable outcome in patients with COVID-19 pneumonia. *Int. J. Environ. Res. Public Health* **18**, 2078 (2021).
45. Szabo, P. A. et al. Analysis of respiratory and systemic immune responses in COVID-19 reveals mechanisms of disease pathogenesis. *MedRxiv Prepr. Serv. Health Sci.* 2020.10.15.20208041 (2020) <https://doi.org/10.1101/2020.10.15.20208041>.
46. Hue, S. et al. Uncontrolled innate and impaired adaptive immune responses in patients with COVID-19 acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* **202**, 1509–1519 (2020).
47. Alay, H. & Laloglu, E. The role of angiotensin-converting enzyme 2 and surfactant protein-D levels in SARS-CoV-2-related lung injury: a prospective, observational, cohort study. *J. Med. Virol.* **93**, 6008–6015 (2021).
48. Lemańska-Perek, A. & Adamik, B. Fibronectin and its soluble EDA-FN isoform as biomarkers for inflammation and sepsis. *Adv. Clin. Exp. Med. Off. Organ Wroclaw Med. Univ.* **28**, 1561–1567 (2019).
49. Beltrán-Camacho, L. et al. The serum of COVID-19 asymptomatic patients up-regulates proteins related to endothelial dysfunction and viral response in circulating angiogenic cells ex-vivo. *Mol. Med.* **28**, 40 (2022).
50. Kosacka, M., Dyla, T., Chaszczewska-Markowska, M., Bogunia-Kubik, K. & Brzecka, A. Decreased thrombospondin-1 and bone morphogenetic protein-4 serum levels as potential indices of advanced stage lung cancer. *J. Clin. Med.* **10**, 3859 (2021).
51. Shahbazi, M. et al. Linkage of lambda interferons in protection against severe COVID-19. *J. Interferon Cytokine Res. Off. J. Int. Soc. Interferon Cytokine Res.* **41**, 149–152 (2021).
52. Plotnikova, M. et al. IFN- $\lambda$ 1 displays various levels of antiviral activity in vitro in a select panel of RNA viruses. *Viruses* **13**, 1602 (2021).
53. Jeon, Y. J. et al. The influence of interferon-lambda on restricting Middle East respiratory syndrome coronavirus replication in the respiratory epithelium. *Antiviral Res.* **180**, 104860 (2020).

### Author contributions

C.S., K.H. and A.G. conceived and supervised the project. A.F., C.D., M.M.A., U.M. conceived and performed the clinical studies, contributed samples, clinical annotation

and supported the data interpretation. N.S., R.G. and K.H. performed the experiments. M.K., F.S., K.H. and C.S. contributed to data analysis, machine learning and data visualization. J.D.H., C.L. and R.S. contributed to planning and provided ideas in order to optimize the described work and interpret the data. K.H., A.F. and N.S. interpreted the data and wrote the original draft of the manuscript. All authors contributed to manuscript writing and approved the final version of the manuscript.

### Competing interests

The authors declare the following competing interests: C.S. and J.D.H. are founders and shareholders of Sciomics GmbH. C.S., K.H., N.S., M.K., A.G. and F.S. are patent inventors who contributed to a patent claiming Covid-19 disease severity biomarkers (WO2022028917). All other authors have no competing interests to declare.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43856-023-00283-z>.

**Correspondence** and requests for materials should be addressed to Christoph Schröder.

**Peer review information** *Communications Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023