



An inflammatory cytokine signature predicts COVID-19 severity and survival

Diane Marie Del Valle^{1,2,3,14}, Seunghee Kim-Schulze^{1,2,3,4,14}, Hsin-Hui Huang^{5,6,7,14}, Noam D. Beckmann⁸, Sharon Nirenberg^{8,9}, Bo Wang¹⁰, Yonit Lavin¹⁰, Talia H. Swartz¹⁰, Deepu Madduri¹⁰, Aryeh Stock¹¹, Thomas U. Marron^{2,3,10}, Hui Xie¹, Manishkumar Patel¹, Kevin Tuballes¹, Oliver Van Oekelen⁸, Adeb Rahman^{1,2,3,8}, Patricia Kovatch^{8,9}, Judith A. Aberg¹⁰, Eric Schadt⁸, Sundar Jagannath¹⁰, Madhu Mazumdar^{5,6,7}, Alexander W. Charney⁸, Adolfo Firpo-Betancourt¹¹, Damodara Rao Mendu¹¹, Jeffrey Jhang¹¹, David Reich¹², Keith Sigel¹⁰, Carlos Cordon-Cardo¹¹, Marc Feldmann¹³, Samir Parekh^{3,4,10}, Miriam Merad^{1,2,3,4,10} and Sacha Gnjjatic^{1,2,3,4,10,11} ✉

Several studies have revealed that the hyper-inflammatory response induced by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a major cause of disease severity and death. However, predictive biomarkers of pathogenic inflammation to help guide targetable immune pathways are critically lacking. We implemented a rapid multiplex cytokine assay to measure serum interleukin (IL)-6, IL-8, tumor necrosis factor (TNF)- α and IL-1 β in hospitalized patients with coronavirus disease 2019 (COVID-19) upon admission to the Mount Sinai Health System in New York. Patients ($n = 1,484$) were followed up to 41 d after admission (median, 8 d), and clinical information, laboratory test results and patient outcomes were collected. We found that high serum IL-6, IL-8 and TNF- α levels at the time of hospitalization were strong and independent predictors of patient survival ($P < 0.0001$, $P = 0.0205$ and $P = 0.0140$, respectively). Notably, when adjusting for disease severity, common laboratory inflammation markers, hypoxia and other vitals, demographics, and a range of comorbidities, IL-6 and TNF- α serum levels remained independent and significant predictors of disease severity and death. These findings were validated in a second cohort of patients ($n = 231$). We propose that serum IL-6 and TNF- α levels should be considered in the management and treatment of patients with COVID-19 to stratify prospective clinical trials, guide resource allocation and inform therapeutic options.

As of late July 2020, COVID-19 disease, caused by SARS-CoV-2 infection, has resulted in more than 15.5 million infections and 634,000 deaths worldwide. A recent study of hospitals in New York City, at the initial epicenter of the COVID-19 pandemic in the United States, reported that, during March 2020, 21% of patients hospitalized with confirmed COVID-19 died¹. These findings are aligned with outcomes observed in the Mount Sinai Health System^{2,3}. There are currently no curative or preventive therapies for COVID-19, highlighting the need to enhance current understanding of SARS-CoV-2 pathogenesis for the rational development of therapeutics.

Recent studies have suggested that, in addition to direct viral damage, uncontrolled inflammation contributes to disease severity in COVID-19 (refs. 4,5). Consistent with this hypothesis, high levels of inflammatory markers, including C-reactive protein (CRP), ferritin and D-dimer, high neutrophil-to-lymphocyte ratio^{6–9} and increased levels of inflammatory cytokines and chemokines^{6,8–11} have been observed in patients with severe diseases. Pathogenic inflammation,

also referred to as cytokine storm, shares similarities with what was previously seen in patients infected with other severe coronaviruses, including SARS-CoV and Middle East respiratory syndrome coronavirus¹², and bears similarities to cytokine release syndrome (CRS) observed in patients with cancer treated with chimeric antigen receptor-modified (CAR) T cells¹³. Tocilizumab, an IL-6 receptor inhibitor, is a US Food and Drug Administration (FDA)-approved treatment for CRS in patients receiving CAR T cells¹⁴. Several single-center studies have used IL-6 inhibitors to treat patients with COVID-19 with some clinical benefits¹⁵ and reported failures¹⁴. Beyond IL-6, several cytokines have been shown to be elevated in CRS and to contribute to tissue damage. TNF- α is important in nearly all acute inflammatory reactions, acting as an amplifier of inflammation. TNF- α blockade has been used to treat more than ten different autoimmune inflammatory diseases, suggesting that this might be a potential therapeutic approach to reduce organ damage in patients with COVID-19 (ref. 16). IL-1 is also a highly active pro-inflammatory cytokine, and monotherapy blocking

¹Human Immune Monitoring Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ²Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ³Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁴Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁵Institute for Health Care Delivery Science, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁶TCI Biostatistics Shared Resource Facility, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁷Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁸Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁹Department of Scientific Computing, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹⁰Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹¹Pathology, Molecular and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹²Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ¹³Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK. ¹⁴These authors contributed equally: Diane Marie Del Valle, Seunghee Kim-Schulze, Hsin-Hui Huang. ✉e-mail: sacha.gnjatic@mssm.edu

IL-1 activity is used to treat inflammatory diseases, including rheumatoid arthritis and inherited auto-inflammatory syndromes, such as cryopyrin-associated syndromes, and has led to sustained reduction in disease severity¹⁷. IL-8 is a potent pro-inflammatory cytokine playing a key role in the recruitment and activation of neutrophils during inflammation¹⁸, and, given the frequent neutrophilia observed in patients infected with SARS-CoV-2, it is possible that IL-8 contributes to COVID-19 pathophysiology.

To mitigate inflammation caused by SARS-CoV-2, immunomodulatory agents, including small molecules and monoclonal antibodies targeting cytokines, have rapidly been entering into clinical trials⁴, and many such FDA-approved agents are already being used routinely in the clinic in an off-label manner. Given the significant side effects associated with the use of these agents, there is an urgent need to identify biomarkers that can accurately predict which patients will deteriorate from an unchecked inflammatory response and help guide rational targeted immunomodulatory therapeutic strategies.

In this study, we asked whether inflammatory cytokine levels can help predict disease course and outcome in patients with COVID-19. To enhance the relevance of the cytokine assays, we focused on four pathogenic cytokines—IL-6, IL-8, TNF- α and IL-1 β —with clinically available or experimental drugs to counteract them. Clinical specimens were analyzed on the ELLA microfluidics platform (see Methods). We selected this platform owing to the rapid turnaround time of assay results (within 3 h of sample collection), making these results potentially actionable.

We followed 1,484 patients hospitalized for suspected or confirmed COVID-19 at the Mount Sinai Health System from the day of hospitalization to the day of discharge or death. We measured serum IL-6, IL-8, TNF- α and IL-1 β levels upon admission and correlated these results with clinical and laboratory markers of disease severity and with disease outcome. We found that elevated IL-6 and TNF- α serum levels at presentation were strong predictors of disease severity and survival, independently of other standard biomarker measurements of laboratory and clinical severity factors. These results suggest that multiplex cytokine profiling could be used to stratify patients and guide resource allocation and prospective interventional studies.

Results

Cohort characteristics and cytokine ranges. We obtained laboratory and health information as part of standard clinical care from 1,484 patients with suspected or confirmed SARS-CoV-2 infection and hospitalized at the Mount Sinai Health System in New York City between March 21 and April 28, 2020, under expedited institutional review board (IRB) approval. Using an emergency use approval from the New York State Department of Health, we implemented the ELLA microfluidics soluble analyte test in the clinical laboratories to measure four inflammatory cytokines known to contribute to pathogenic inflammation in CAR T cell-associated CRS—IL-6, IL-8, TNF- α and IL-1 β —and assessed their correlation with severity and survival. Of the patients tested, 1,257 had a documented positive or presumptive positive SARS-CoV-2 polymerase chain reaction (PCR) test, whereas the remaining 167 could not be confirmed.

A total of 1,953 specimens were analyzed to quantify circulating IL-6, IL-8, TNF- α and IL-1 β serum levels using the ELLA rapid detection enzyme-linked immunosorbent assay (ELISA) microfluidics platform (Methods and Extended Data Fig. 1a–e). In most of the 1,484 patients accrued, samples were collected once, typically upon admission to the hospital (median, 1.2 d; interquartile range (IQR), 0.7–3.0 d). A subset of patients ($n = 244$) had cytokine measurements performed more than once after admission, although, for all prognostic analyses, only the first available test was used. For the entire cohort, the median time available from first cytokine test to last follow-up (that is, date of discharge, date of death or date

Table 1 | Patient characteristics

| | Patients with SARS-Cov-2 confirmed by PCR | Patients without SARS-Cov-2 confirmed by PCR |
|---|---|--|
| Median age (IQR)—years | 63 (53–72) | 60 (49–73) |
| Male | 787/1,309 (60.1%) | 90/167 (53.9%) |
| Race/ethnicity—Hispanic | 577/1,268 (45.5%) | 62/167 (37.1%) |
| Race/ethnicity—African American | 278/1,268 (21.9%) | 46/167 (27.5%) |
| Race/ethnicity—White | 277/1,268 (21.8%) | 43/167 (25.7%) |
| Race/ethnicity—Asian | 73/1,268 (5.8%) | 5/167 (3.0%) |
| Race/ethnicity—Other | 63/1,268 (5.0%) | 11/167 (6.6%) |
| Obesity (BMI ≥ 30) | 465/1,176 (39.5%) | 34/149 (22.8%) ^a |
| Comorbidities—hypertension | 420/1,268 (33.1%) | 67/167 (40.1%) |
| Comorbidities—diabetes | 293/1,268 (23.1%) | 34/167 (20.4%) |
| Comorbidities—CKD | 167/1,268 (13.2%) | 27/167 (16.2%) |
| Comorbidities—cancer (active) | 147/1,267 (11.6%) | 37/166 (22.3%) |
| Comorbidities—atrial fibrillation | 123/1,267 (9.7%) | 9/167 (5.4%) ^a |
| Comorbidities—CHF | 69/1,268 (5.4%) | 18/167 (10.8%) ^a |
| Comorbidities—asthma | 72/1,268 (5.7%) | 13/167 (7.8%) |
| Comorbidities—COPD | 44/1,268 (3.5%) | 18/167 (10.8%) |
| Comorbidities—sleep apnea | 50/1,268 (3.9%) | 6/167 (3.6%) |
| Smoking—current | 55/965 (5.7%) | 27/143 (18.9%) ^a |
| Smoking—history | 293/965 (30.4%) | 53/143 (37.1%) |
| SARS-CoV-2 PCR detected | 1,257/1,257 (100.0%) | 0/167 (0.0%) ^a |
| SARS-CoV-2 Ab detected | 65/73 (89.0%) | 3/5 (60.0%) |
| Maximum severity score ^b —mild/moderate | 505/1,168 (43.2%) | 99/150 (66.0%) ^a |
| Maximum severity score ^b —severe | 285/1,168 (24.4%) | 22/150 (14.7%) ^a |
| Maximum severity score ^b —severe with end organ damage | 378/1,168 (32.4%) | 29/150 (19.3%) ^a |
| SOFA score = 0 | 309/870 (35.5%) | 42/119 (35.3%) |
| SOFA score = 1 | 163/870 (18.7%) | 24/119 (20.2%) |
| SOFA score >1 (median = 4) | 398/870 (45.7%) | 53/119 (44.5%) |
| Acute respiratory distress syndrome | 199/1,267 (15.7%) | 11/167 (6.6%) ^a |
| Died within follow-up period | 269/1,317 (20.4%) | 13/167 (7.8%) ^a |

^aSignificantly different between PCR-positive and PCR-negative subsets by Fisher's exact test or chi-squared test. ^bSeverity score, described in Methods, is a composite measurement for COVID-19 established by Mount Sinai infectious disease and pulmonology teams based on criteria used in the literature and includes CrCl, ALT levels and use and type of respirators/ventilators and vasopressors (see Methods). Severity score indicated here is the maximum achieved throughout the observation period, but, for all other analyses, the severity score at the time of the first measurement was used. Denominator indicates number of patients with available information. CKD, chronic kidney disease; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; Ab, antibody.

still in hospital, whichever was latest) was 8 d (IQR, 3.1–16.0 d, up to 41 d). Patient characteristics are listed in Table 1. As references, and to serve as controls, cytokine measurements collected before the launch of this study were performed in healthy donors and in

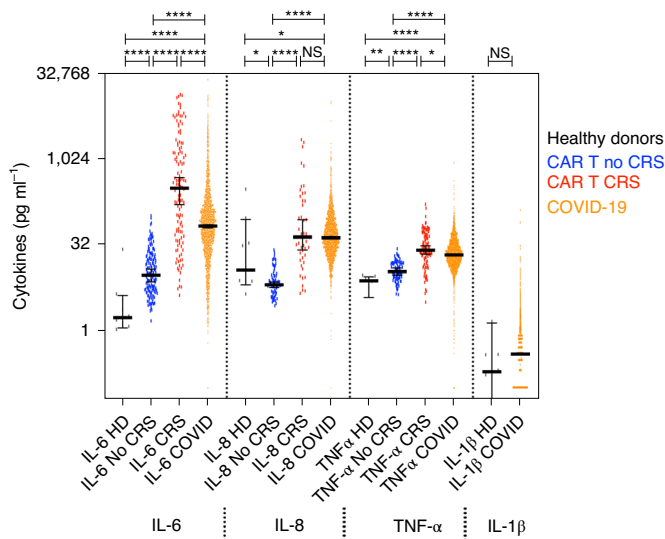


Fig. 1 | Range of measured cytokines. Detection range of cytokines in all tested serum samples from patients with COVID-19 hospitalized at the Mount Sinai Health System (orange, $n=1,959$), in comparison with serum samples from healthy donors (black, $n=9$) and plasma samples from patients with multiple myeloma prior to (blue, $n=151$) and during (red, $n=121$) CRS induced by CAR T cell therapy. Heavy bars indicate median, and error bars represent 95% CI, each value indicated by a dot. Pairwise comparisons by the two-sided Mann-Whitney t -test show significantly higher levels of IL-6, IL-8 and TNF- α in COVID-19 samples compared to samples from healthy donors of patients with non-CRS cancer ($****P < 0.0001$, $***P < 0.001$, $**P < 0.01$ and $*P < 0.05$; NS, not significant). Median, mean and range are shown in Extended Data Fig. 1d (error band indicates the median with 95% CI). HD, hemodialysis.

patients with cancer who either developed or did not develop CRS after CAR T cell therapies^{19,20}.

We found that IL-6 ($P < 0.0001$), IL-8 ($P < 0.0001$) and TNF- α ($P < 0.0001$) were significantly elevated in COVID-19 serum compared to healthy donor serum or plasma isolated from CAR T cell-treated patients with no CRS (Fig. 1). The four cytokines assessed had different detection ranges, with IL-6 having the most dynamic profile, followed by IL-8 and TNF- α (Fig. 1 and Extended Data Fig. 1d). In line with previous reports, IL-1 β levels were mostly low or at the limit of detection of 0.1 pg ml⁻¹, even though the assay was able to detect various levels of recombinant control cytokines (Extended Data Fig. 1b). The vast majority of patients, therefore, presented with elevated cytokines or cytokine storm, but, in contrast to the coordinated increase in cytokines during CAR T CRS (average Spearman's $r=0.6$), cytokine levels were not as highly correlated with each other in COVID-19 samples (average Spearman's $r=0.4$), suggesting differential patterns of cytokine expression and potentially distinct clinical presentations based on the relative profile of each independent cytokine (Extended Data Fig. 1e,f). Because more than 70% of samples analyzed for each cytokine in COVID-19 fell within the CRS range based on our post-CAR-T-defined cut-offs, and because we did not have an established cutoff for IL-1 β , we decided to separate high versus low values using a cutoff above the median for each cytokine in patients with COVID-19. After empirical testing as described in the Methods, the cutoffs chosen for further statistical analyses were more than 70 pg ml⁻¹ for IL-6, more than 50 pg ml⁻¹ for IL-8, more than 35 pg ml⁻¹ for TNF- α and more than 0.5 pg ml⁻¹ for IL-1 β .

Association with demographics and comorbidities. We used the first available cytokine measurement in each patient to measure

correlations with demographics and comorbidities. We hypothesized that cytokines are elevated in patients with COVID-19 compared to healthy donors and non-CRS CAR-T-treated patients owing to SARS-CoV-2 infection. Of the 1,484 patients hospitalized with COVID-19 symptoms, 11.7% tested negative for SARS-CoV-2 by PCR and, therefore, were excluded from further univariate analyses. It should be noted that there might have been false-negative tests for SARS-CoV-2 viral detection based on subsequent tests demonstrating antibodies to SARS-CoV-2 S spike protein in three of five patients who tested negative for SARS-CoV-2 by PCR. Despite similar comorbidities, cytokine levels in this subset of patients were significantly lower compared to patients who tested positive for SARS-CoV-2 infection (Fig. 2a). Of the remaining patients who tested positive for SARS-CoV-2 by PCR, 1,097 had complete information for demographics and comorbidities.

Men had significantly higher levels of IL-6 than women ($P < 0.0001$), but no sex differences were observed for the other three cytokines (Fig. 2b). With increased age brackets (<50, 50–70 and >70 years old), levels of IL-6, IL-8 and TNF- α increased (Fig. 2b), and the same was observed for age when assessed as a continuous variable. There was no association of any cytokine measured with body mass index (BMI). Smoking and race/ethnicity showed weak but significant univariate associations with IL-6, IL-1 β and/or TNF- α , which were not confirmed after adjusting for the other covariates, except for IL-1 β and TNF- α , which remained significantly higher when comparing Hispanics to African Americans.

We then assessed whether cytokine levels were associated with comorbidities listed in Table 1. We found that TNF- α and IL-8 were significantly increased in patients with chronic kidney disease (CKD), diabetes and hypertension, whereas TNF- α was also increased in patients with congestive heart failure (CHF), based on univariate analyses. IL-6 and IL-8 were elevated in patients with a history of atrial fibrillation. No associations were found between cytokines and active cancer, asthma, chronic obstructive pulmonary disease (COPD), human immunodeficiency virus (HIV) and sleep apnea.

Using multivariable regression models, we confirmed that CKD was the only comorbidity significantly associated with elevated cytokine levels, whereas elevated TNF- α in patients with diabetes and hypertension were explained by other variables. Of demographic variables, age and sex (for IL-6) remained significantly associated with cytokine levels as seen in univariate analyses. Therefore, we included demographics and comorbidities as confounding variables in subsequent analyses. Cytokine levels, as measured by ELLA, were not significantly affected by timing of testing in relation to hospital admission. Therefore, this time difference was not considered as a potential confounder.

Association between cytokines and risk of death. Next, we considered factors affecting survival defined as time to death and censored regardless of cytokines in the overall cohort with univariate Kaplan–Meier analyses. We found that only age and CKD were significantly associated with increased risk of death from COVID-19. We evaluated whether cytokines could distinguish patients based on overall survival and disease severity after COVID-19 hospitalization. Stratifying patients by cytokine levels of high versus low using the cutoffs described in the statistical analysis section, we found that each cytokine could predict the overall survival of patients, based on the first available measurement after hospital admission. Each cytokine was independently predictive of overall survival, after adjusting for demographics and comorbidities—that is, sex, age, race/ethnicity, smoking, CKD, hypertension, asthma and CHF (Fig. 3).

When considering all cytokines together in the model, all but IL-1 β remained significant, even after adjustment for demographics and comorbidities ($n=1,097$). This confirmed the relative independence of each cytokine tested, with only age (50–70 versus

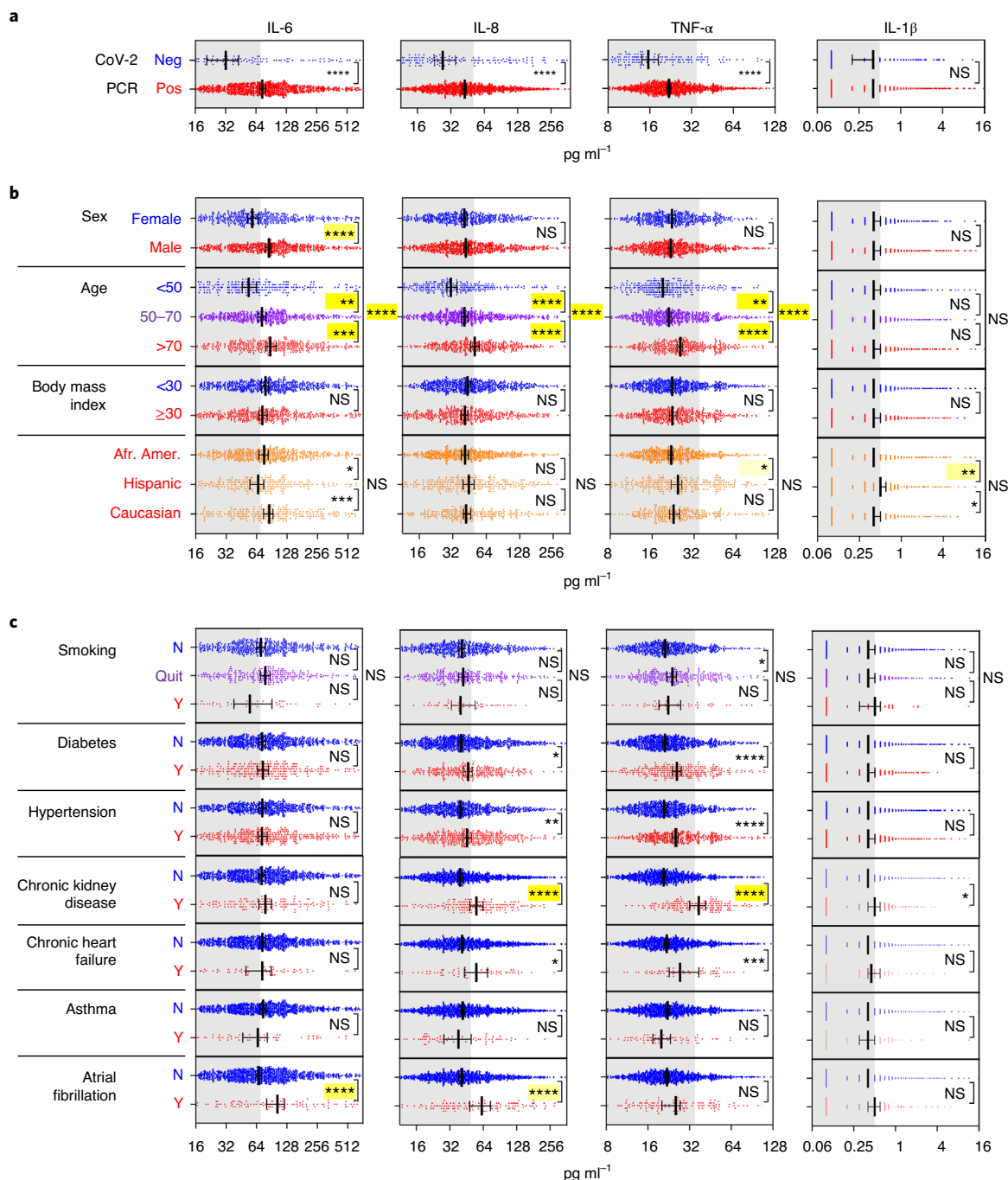


Fig. 2 | Cytokine levels by PCR status, demographics and comorbidity. Cytokine levels observed in relation to **a**, SARS-CoV-2 PCR status (negative indicates patients with COVID-19-like respiratory symptoms with a negative SARS-CoV-2 PCR test) ($n=1,422$ independent patient samples); **b**, demographics (excluding PCR negative, with data available for sex, age, BMI and race/ethnicity for 1,298, 1,307, 1,174 and 1,131 patients, respectively); and **c**, comorbidities (excluding PCR negative, data available for smoking and comorbidity diagnoses, respectively, for 964 and 1,266 individual patients). Scatter plots indicating individual measurements (dots); thick line is median; error bars representing 95% CI; and statistical analyses by two-sided Mann-Whitney univariate t -test (**** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$ and * $P < 0.05$; NS, not significant). Not shown here are COPD, HIV, sleep apnea and active cancer, which did not show any significant difference for cytokine levels. In yellow highlights are the statistical values that were still significant after adjustment of all demographic and comorbidity variables, with shade of yellow indicating adjusted P value (light: *, mid: **, high: *** and saturated: ****). Gray area indicates cytokine levels below the respective cutoff.

<70 years, hazard ratio (HR) = 2.09 (1.25–3.49); >70 versus <50 years, HR = 3.76 (2.24–6.33)), IL-6 (HR = 2.23 (1.61–3.09)), IL-8 (HR = 1.41 (1.05–1.89)) and TNF- α (HR = 1.50 (1.09–2.07)) remaining significantly associated with decreased survival after adjustments ($P=0.0049$, $P < 0.0001$, $P=0.0205$ and $P=0.0140$,

respectively). Internal validation for this model achieved an uncorrected concordance index of 0.738, a ten-fold coefficient of variation (CV) concordance index of 0.705 and a bootstrap-corrected concordance index of 0.716. As additional validation, we also performed this analysis using a competing risk model, in which

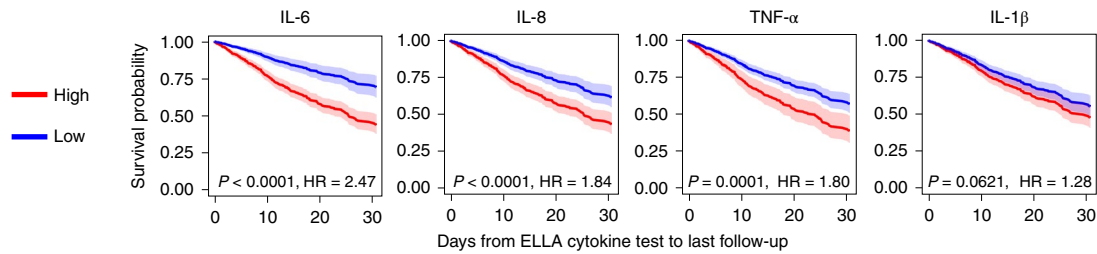


Fig. 3 | Cytokine levels and survival. Survival curves based on each cytokine measured, after multiple variable adjustments for sex, age, race/ethnicity, smoking, CKD, hypertension, asthma and CHF ($n=1,246$). Cox regression model showing overall survival with CIs for each cytokine based on time from ELLA cytokine test to last follow-up date (discharge, death or still in hospital, whichever comes last), with significance indicated by P value and HR. There was worse survival if cytokines were high (red, above cutoffs of 70 pg ml^{-1} for IL-6, 50 pg ml^{-1} for IL-8, 35 pg ml^{-1} for TNF- α and 0.5 pg ml^{-1} for IL-1 β) versus low (blue, below cutoffs). Each line indicates the predicted survival probability over follow-up time, with the error band indicating the corresponding two-sided 95% CI.

patients discharged alive were considered competing events, and patients in hospital were censored, and found the same conclusion, where high IL-6, IL-8 and TNF- α remained significantly associated with worse outcome regardless of demographics and comorbidities (Supplementary Table 1). We used the competing risk model in the next analysis.

Using cytokines to complement risk stratification. Next, we asked whether cytokines were of value for risk stratification and survival, independent of known laboratory and clinical severity metrics (that is, temperature, O_2 saturation, respiratory rate and severity score as defined in Methods and Extended Data Fig. 2). We first tested whether the four tested cytokine levels were associated with known inflammation markers CRP, D-dimer and ferritin and found strong correlations in all cytokines with each measurement, with IL-6 and IL-1 β additionally associated with fever (Fig. 4a). In addition, IL-6 and IL-8 levels were closely correlated with severity scale (moderate, severe and severe with end organ damage), which takes into account lung imaging, creatinine clearance (CrCl), vasoactives and use of ventilation, whereas TNF- α did not distinguish moderate versus severe COVID-19 presentation or use of mechanical ventilation, but, instead, was only increased with end organ damage. Looking at the predictive value of cytokines on survival after adjusting for levels of CRP, D-dimer, ferritin and all comorbidities, IL-6 and IL-8 remained independently predictive of survival, therefore showing additive value to these known markers (Supplementary Table 2). When including additional severity metrics, including severity scoring, IL-8 was no longer predictive of survival, likely because these added parameters were stronger factors for the competing risk model (Supplementary Table 3).

We then investigated correlations of cytokines with an additional series of well-established markers of inflammation, renal function, myocardial strain and respiratory distress for their effect on survival within this cohort. Using unsupervised analyses, neutrophils, white blood cells, CRP, ferritin, D-dimer, lactate dehydrogenase and low O_2 saturation co-clustered with all cytokines except TNF- α , which was more closely correlated with markers of tissue damage such as creatinine (Extended Data Fig. 3). Selecting the most informative variables using a backward elimination process to define each of the available measurements to be used as confounding factors in a competing risk regression analysis of survival along with the cytokines, we found severity score, O_2 saturation, platelets, low albumin, systolic blood pressure, D-dimer, albumin, calcium, chloride and platelet count remaining. Remarkably, even when using these measurements as variables to adjust when assessing the predictive value of cytokines on survival in the competing risk regression analysis ($n=802$), we found that IL-6 and TNF- α remained significantly associated with a worse prognosis (Fig. 4b). Internal validation with

this model achieved an uncorrected concordance index of 0.794 and a corrected index of 0.764 (ten-fold CV and bootstrap). In a subset of patients ($n=663$), Sequential Organ Failure Assessment (SOFA) severity scale scores were also available, and we confirmed that IL-6 (HR = 2.9, $P < 0.0001$), IL-8 (HR = 1.6, $P = 0.04$) and TNF- α (HR = 1.6, $P = 0.03$) were associated with poor survival, after adjusting for all most informative variables above, including increased SOFA severity (treated either as a continuous variable or as SOFA score of ≤ 1 versus > 1).

Finally, we applied the survival models from our analysis (the primary model: cytokines, demographics and comorbidities; the secondary model: the primary model plus the markers of inflammation, renal function, myocardial strain and respiratory distress) to an independent validation cohort of 231 hospitalized patients who tested positive for SARS-CoV-2 by PCR collected between April 22 and June 16, 2020, with available cytokine, demographics, comorbidity and laboratory data. The area under the receiver operating characteristic curve (AUC) plots showed that the primary model performs well between days 3 and 31, during which the AUC ranged from 0.65 to 0.76 (Extended Data Fig. 4a). The secondary model had somewhat higher AUC, ranging from 0.70 to 0.88 (Extended Data Fig. 4d). The integrated AUCs of the two models were 0.68 and 0.74, respectively. The actual and the predicted survival probabilities were similar until day 20, after which the two curves separated (Extended Data Fig. 4b,e). The distributions of the prognostic indices were not significantly different between the original and validation cohorts for the primary model ($P = 0.11$) and the secondary model ($P = 0.06$) (Extended Data Fig. 4c,f).

Therefore, we conclude that IL-6 and TNF- α are independently predictive of patient outcomes in terms of both disease severity and survival (Supplementary Table 4). Even after stratifying for risk factors with the strongest P value—that is, severity score, O_2 saturation and age—IL-6 and TNF- α remained independently predictive of survival, with IL-8 also reaching significance (Fig. 4c and Supplementary Table 5).

Effect of medication and treatment on cytokine levels. Although our data do not demonstrate a causative role for IL-6 and TNF- α in disease outcome, we wanted to shed light on the effects of various treatments on measured cytokines as potential mitigation strategies should there be a pathogenic effect from these inflammatory agents. From a subset of 244 patients with more than one ELLA cytokine assay performed, and by mapping time from treatment start to first ELLA test, we were able to assess the effects of various treatments and experimental drugs on cytokine levels (Fig. 5a). Our analysis of a subset of patients with progressive respiratory failure and marked systemic inflammation who received off-label treatment with the anti-IL-6 receptor monoclonal antibody tocilizumab

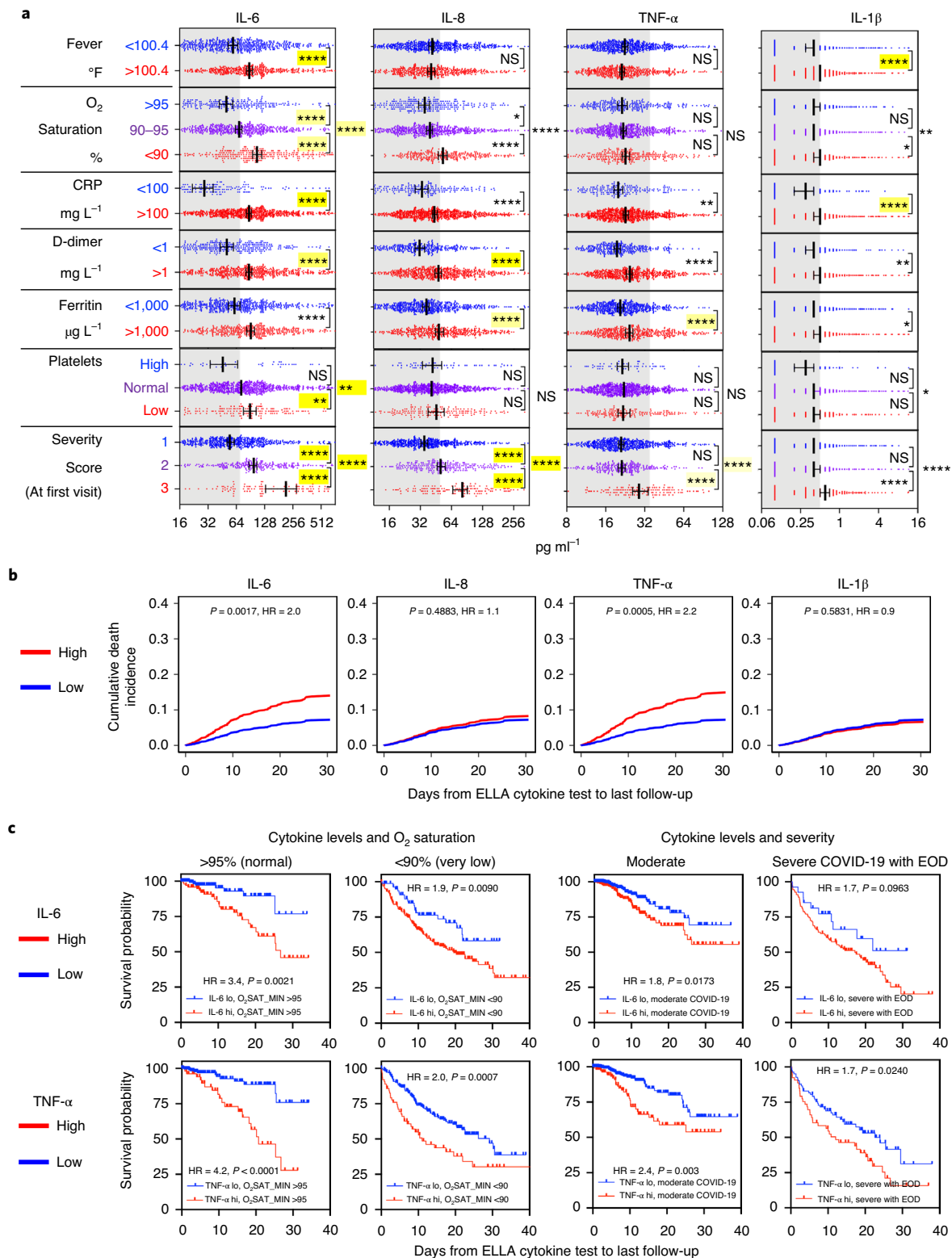


Fig. 4 | Cytokine levels correlate with severity and independently predict survival. Correlation of cytokine levels with established inflammatory and severity measurements. **a**, Correlation of each cytokine with each metric ($n=1,106$ for fever, $n=1,112$ for O₂ saturation, $n=1,023$ for CRP, $n=926$ for D-dimer, $n=1,017$ for ferritin, $n=1,038$ for platelets and $n=1,023$ for disease severity score), using the same univariate and multivariate analyses as in the Fig. 2 legend. Error bar indicates the median \pm 95% CI. **b**, Competing risk analysis ($n=671$) showing survival differences by IL-6 and TNF- α levels, after adjusting the following variables: IL-6, IL-8, TNF- α , IL-1 β , age, sex, race/ethnicity, smoking status, asthma, atrial fibrillation, cancer, CHF, CKD, COPD, diabetes, hypertension, sleep apnea, severity, systolic blood pressure max, O₂ saturation min, D-dimer, albumin, calcium, chloride and platelet count. **c**, Kaplan-Meier univariate analyses of survival by IL-6 and TNF- α levels in patients with normal ($n=257$), low ($n=258$) or very low ($n=287$) O₂ saturation, or in patients with moderate ($n=588$) versus severe COVID-19 with end organ damage ($n=136$), as measured at the first available test. EOD, end organ damage.

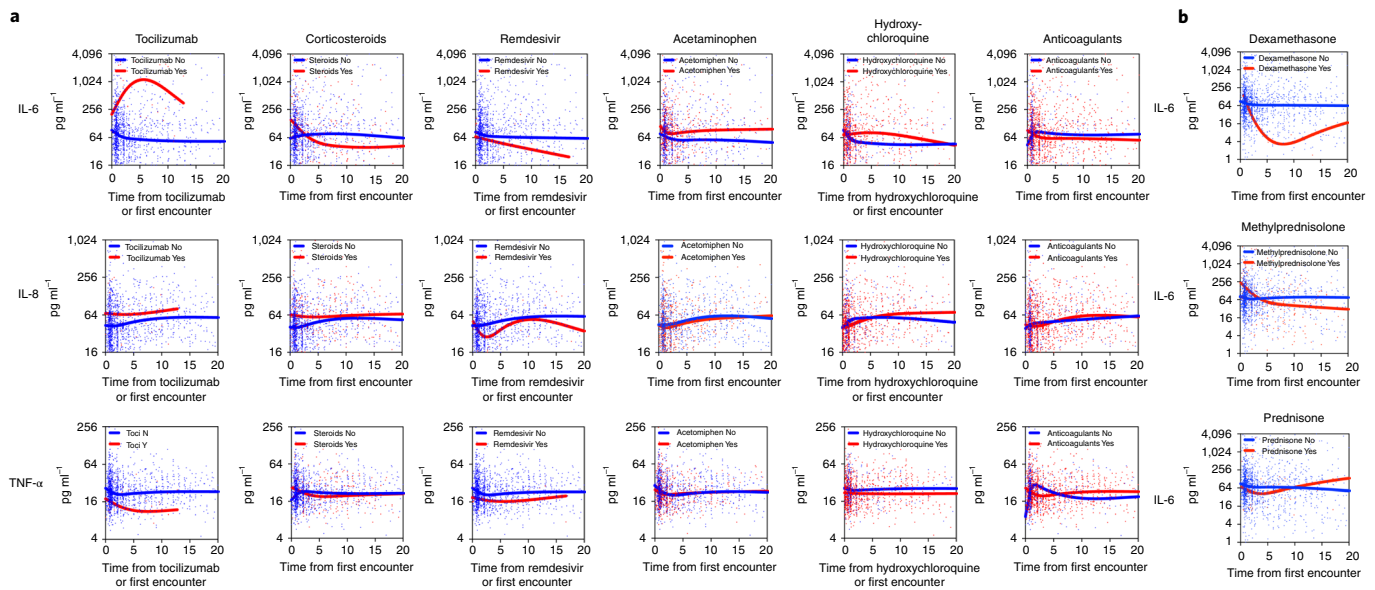


Fig. 5 | Treatment effect on cytokine levels. **a**, Effect of treatments on IL-6 (top row), IL-8 (middle row) and TNF- α (bottom row). Lines (in red: with indicated treatment; in blue: without indicated treatment) represent the best fit curve by smoothed spline of the longitudinal and unique time point distribution of each cytokine level based on time from either first encounter or treatment start. Of 1,670 samples representing various time points of 1,315 patients with available information, the number of those from patients who received tocilizumab, corticosteroids (any of prednisone, methylprednisolone or dexamethasone), remdesivir, acetaminophen, hydroxychloroquine and/or anticoagulants (apixaban, enoxaparin, heparin or rivaroxaban) was 73, 305, 76, 620, 1,333, and 1,113, respectively. **b**, Effect of different corticosteroids on IL-6.

showed that these patients started with elevated IL-6 levels and then had a transient increase in serum IL-6, which has previously been explained by disrupted clearance after drug saturation of the IL-6 receptor²¹. This transient elevation was observed only for IL-6, not IL-8, whereas TNF- α appeared to gradually decrease after therapy. Patients treated with corticosteroids and remdesivir showed, respectively, a rapid and gradual reduction in IL-6 over time compared to patients who did not receive these drugs, but we observed no effect on TNF- α . Hydroxychloroquine, acetaminophen or anti-coagulants did not clearly appear to alter cytokine levels. Of corticosteroids, dexamethasone had the highest reduction effect on IL-6 (Fig. 5b), potentially supporting findings from the recent RECOVERY trial showing clinical benefit from this drug in hospitalized patients with severe disease²².

Discussion

We aimed to understand the role of inflammatory cytokines on COVID-19 disease course and outcome. We established a rapid multiplex cytokine test to measure IL-6, TNF- α and IL-1 β , as known markers of inflammation and organ damage, along with CXCL8/IL-8 because of its potent role in the recruitment and activation of neutrophils, commonly elevated in patients with COVID-19 (ref. ²³). Notably, drugs blocking these cytokines are either FDA approved or in clinical trials. Studying over 1,400 hospitalized patients in a month, we established that COVID-19 is associated with high levels of all four cytokines at presentation. Importantly, our observations indicate that cytokine patterns are predictive of COVID-19 survival and mortality, independently of demographics and comorbidities, but also of standard clinical biomarkers of disease severity, including laboratory and clinical factors. A model based on these observations was confirmed in a validation cohort of another 231 patients. We found that IL-6 was one of the most robust prognostic markers of survival, eclipsing or outperforming CRP, D-dimer and ferritin after adjusting for the demographic features and comorbidities. It remained independently associated with severity and predictive of outcome when including information about ventilation and end

organ damage. Furthermore, elevated TNF- α , known to contribute to organ damage, was also a strong predictor of poor outcome even after adjusting for other risk factors such as age, sex, hypoxia, disease severity scoring based on clinical assessment and IL-6. Our cytokine panel also included IL-8, which showed association with survival time, even though it was eclipsed by other severity factors after multivariate adjustments, and IL-1 β , which was poorly detected and, as a result, had only marginal predictive value. Although classic markers used routinely to determine inflammation and severity were still useful to stratify patients on their own, when combined in multivariate analyses, many were no longer significant, likely due to collinearity, whereas IL-6 and TNF- α remained independently predictive of outcome. Both overall survival and competing risk models used here consistently showed the significant prognostic value of TNF- α and IL-6 when all tested cytokines were in the model, along with demographics, comorbidities and other clinical and laboratory measurements, highlighting the robustness of our findings. Notably, the COVID-19-related cytokine response was quite distinct from the traditional cytokine storm associated with sepsis and CAR T cells, with sustained elevated cytokine levels over days and weeks, and relative absence of coordination between cytokines. This raises the possibility of mitigation strategies with anti-cytokine treatments, although which one(s) and the window of opportunity for their use remain to be established. Guiding such therapies based on mechanistic association with cytokine levels could provide a rational approach.

Trials to block IL-6 signaling with already FDA-approved drugs have been launched across the world, and some clinical benefits have been seen in a subset of patients in small, single-center, observational studies^{15,24}. In contrast, interim analysis of randomized trials with the anti-IL-6 receptor monoclonal antibody sarilumab versus placebo identified potential benefit only in patients with severe but not moderate disease (<https://investor.regeneron.com/news-releases/news-release-details/regeneron-and-sanofi-provide-update-us-phase-23-adaptive>). There are no available data correlating levels of IL-6 and response

to treatment, and none of the current studies has used cytokine profiling as part of its inclusion criteria. It is possible that patients with moderate disease and high IL-6 levels will benefit the most from cytokine blockade. Additionally, IL-6 reduction observed in patients treated with dexamethasone might be a mechanism underlying the efficacy of this treatment²⁵. There is also a need to evaluate the effect of anti-TNF- α therapy on its own in COVID-19. Because IL-6 and TNF- α appear to be independent variables, studies with a combination regimen blocking both cytokines would be important to consider for added clinical efficacy.

Early cytokine measurements are reliable predictors of outcome and, therefore, raise the critical importance of using serum cytokine levels for treatment decisions. The predictive value of these cytokines might help inform therapeutic interventions to determine which individuals are likely to develop respiratory failure, end organ damage and death and to select optimal trial designs to disrupt the underlying inflammatory milieu. A prediction model built on cytokine levels early in disease might serve to inform healthcare allocation and prioritization of individuals at highest risk.

Although confirmed in our validation cohort, the predictive value of IL-6 and TNF- α should also be assessed in a prospective manner, where more control over data collection can be applied. Although the focus of this study was on only four cytokines chosen for their known inflammatory or pathogenic properties, additional soluble analytes will likely be useful to consider to refine the survival predictive model. Our current efforts are to build such a predictive model that will make use of a prospectively collected cohort where we will leverage high-dimensional assays such as Olink proximity extension assay and the SomaLogic aptamer platform²⁶, which can measure hundreds to thousands of soluble analytes from serum or plasma. The most informative dimensions from these assays could then be carried back into the rapid 4-8 plex ELLA cytokine detection system for clinical decision-making, in addition to IL-6 and TNF- α . We think that these practices will bring cytokine measurements to standard of care in prognosticating and monitoring patients with COVID-19.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-1051-9>.

Received: 28 May 2020; Accepted: 4 August 2020;

Published online: 24 August 2020

References

- Richardson, S. et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. *JAMA* **323**, 2052–2059 (2020).
- Paranjpe, I. et al. Clinical characteristics of hospitalized COVID-19 patients in New York City. Preprint at <https://doi.org/10.1101/2020.04.19.20062117> (2020).
- Wang, B. et al. A tertiary center experience of multiple myeloma patients with COVID-19: lessons learned and the path forward. *J. Hematol. Oncol.* **13**, 94 (2020).
- Merad, M. & Martin, J. C. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nat. Rev. Immunol.* **20**, 355–362 (2020).
- Vabret, N. et al. Advancing scientific knowledge in times of pandemics. *Nat. Rev. Immunol.* **20**, 338 (2020).
- Mehta, P. et al. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet* **395**, 1033–1034 (2020).
- Ruan, Q., Yang, K., Wang, W., Jiang, L. & Song, J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med.* **46**, 846–848 (2020).
- Wu, C. et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern. Med.* <https://doi.org/10.1001/jamainternmed.2020.0994> (2020).
- Herold, T. et al. Elevated levels of IL-6 and CRP predict the need for mechanical ventilation in COVID-19. *J. Allergy Clin. Immunol.* **146**, 128–136 (2020).
- Yang, Y. et al. Exuberant elevation of IP-10, MCP-3 and IL-1ra during SARS-CoV-2 infection is associated with disease severity and fatal outcome. Preprint at <https://doi.org/10.1101/2020.03.02.20029975> (2020).
- Zhang X. et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature* **583**, 437–440 (2020).
- Fehr, A. R., Channappanavar, R. & Perlman, S. Middle East respiratory syndrome: emergence of a pathogenic human coronavirus. *Annu. Rev. Med.* **68**, 387–399 (2017).
- Moore, J. B. & June, C. H. Cytokine release syndrome in severe COVID-19. *Science* **368**, 473–474 (2020).
- Davila, M. L. et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia. *Sci. Transl. Med.* **6**, 224ra225 (2014).
- Xu, X. et al. Effective treatment of severe COVID-19 patients with tocilizumab. *Proc. Natl Acad. Sci. USA* **117**, 10970–10975 (2020).
- Feldmann, M. et al. Trials of anti-tumour necrosis factor therapy for COVID-19 are urgently needed. *Lancet* **395**, 1407–1409 (2020).
- Dinareello, C. A., Simon, A. & van der Meer, J. W. Treating inflammation by blocking interleukin-1 in a broad spectrum of diseases. *Nat. Rev. Drug Discov.* **11**, 633–652 (2012).
- Baggiolini, M., Walz, A. & Kunkel, S. L. Neutrophil-activating peptide-1/interleukin 8, a novel cytokine that activates neutrophils. *J. Clin. Invest.* **84**, 1045–1049 (1989).
- Jatiani, S. S. et al. Myeloma CAR-T CRS management with IL-1R antagonist anakinra. *Clin. Lymphoma Myeloma Leuk.* <https://doi.org/10.1016/j.clml.2020.04.020> (2020).
- Madduri, D., Dhodapkar, M. V., Lonial, S., Jagannath, S. & Cho, H. J. SOHO State of the Art updates and next questions: T-cell-directed immune therapies for multiple myeloma: chimeric antigen receptor-modified T cells and bispecific T-cell-engaging agents. *Clin. Lymphoma Myeloma Leuk.* **19**, 537–544 (2019).
- Nishimoto, N. et al. Mechanisms and pathologic significances in increase in serum interleukin-6 (IL-6) and soluble IL-6 receptor after administration of an anti-IL-6 receptor antibody, tocilizumab, in patients with rheumatoid arthritis and Castleman disease. *Blood* **112**, 3959–3964 (2008).
- The RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with COVID-19 — preliminary report. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2021436> (2020).
- Liu, J. et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. *EBioMedicine* **55**, 102763 (2020).
- Gritti, G. et al. Use of siltuximab in patients with COVID-19 pneumonia requiring ventilatory support. Preprint at <https://doi.org/10.1101/2020.04.01.20048561> (2020).
- Ledford, H. Coronavirus breakthrough: dexamethasone is first drug shown to save lives. *Nature* **582**, 469 (2020).
- Rohloff, J. C. et al. Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids* **3**, e201 (2014).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

ELLA cytokine test. The ELLA platform is a rapid cytokine detection system based on four parallel singleplex microfluidics ELISA assays run in triplicate within cartridges following the manufacturer's instructions. We first validated IL-6, IL-8 and TNF- α detection by ELLA at the Mount Sinai Human Immune Monitoring Center using plasma from multiple patients with myeloma who were undergoing immunotherapies such as CAR T cells and bispecific antibodies, known to elicit cytokine release storm. Analytical validation (Extended Data Fig. 1a–c) was performed using both reference cytokine controls and biological replicates across different lots of cartridges. The reproducibility was greater than 95%, with an intra-assay CV of 0.8%, an inter-assay CV of 0.4–0.8% for analytes in the high detection range (>250 pg ml $^{-1}$) and a CV of 2.6–4.2% for analytes in the lowest detection range (5–50 pg ml $^{-1}$). Serum and plasma appeared to be equivalent for detection of these cytokines. In March 2020, as the number of COVID-19 cases was increasing in New York City, we transferred the ELLA methodology to the Mount Sinai Hospital Center for Clinical Laboratories, which allowed the ELLA cytokine test to be coded into our electronic health record ordering system as part of a COVID-19 diagnostic panel.

Patient information and data source. This research was reviewed and approved by the Human Research Protection Program at the Icahn School of Medicine at Mount Sinai (ISMMS). The Program for the Protection of Human Subjects is a key component of ISMMS' efforts to ensure human subject protections. It supports our researchers in assuring the ethical conduct of research and compliance with federal, state and institutional regulations and provides a professional office staff to assist investigators, participants and five IRBs. A waiver of informed consent was obtained to query the patient electronic health records. Samples for the RT-PCR SARS-CoV-2 lab test were collected via nasopharyngeal or oropharyngeal swab at one of 53 different Mount Sinai locations, representing outpatient, urgent care, emergency and inpatient facilities. Blood specimens for ELLA were collected via venipuncture within the Mount Sinai Health System. All specimens and imaging were collected as part of standard of care.

Between March 21 and April 28, 2020, 1,484 patients hospitalized with suspicion of COVID-19 were tested for SARS-CoV-2 viral infection status by PCR and for the ELLA cytokine panel, and routine laboratory measurements and blood counts were obtained as part of standard medical care. For validation purposes, we also obtained data from an independent cohort of clinically annotated SARS-CoV-2 PCR-positive hospitalized patients at Mount Sinai in whom cytokine testing was performed between April 22 and June 16, 2020 ($n=231$; median follow-up, 11.6 d, up to 53 d). Patients were identified by querying the pathology department electronic database for individuals with both SARS-CoV-2 PCR-based testing and ELLA cytokine panel. Cytokine data were obtained from pathology department electronic databases, and clinical and demographic data were supplemented with information from the Mount Sinai Data Warehouse.

A list of medical record numbers for patients who had both a SARS-CoV-2 PCR result and an ELLA cytokine panel result in the pathology department electronic database was provided to the Mount Sinai Data Warehouse. Subsequently, demographic and clinical data were extracted from the Epic electronic health record for the identified patients using the Epic Hyperspace (August 2019), Epic Clarity (February 2020) and Epic Caboodle (February 2020) databases via connecting to Oracle (18c Enterprise Edition Release 18.0.0.0.0) and SQL server (Microsoft SQL Server 2016 (SP2-CU11) (KB4527378) - 13.0.5598.27 (X64)) databases, respectively. Additional data elements included lab results, vital signs, O $_2$ therapy, radiology reports for chest imaging, diagnostic outcomes and medications. Data were merged from the various data sources using R version 3.6.1. Large tables were read-in and written using the R packages tidyverse (v. 1.3.0), reshape2 (v. 1.4.4) and readxl (v. 1.3.1).

Clinical follow-up data were collected up to May 7, 2020, for the main cohort and to June 23, 2020, for the validation cohort. Two investigators (D.M.D.V. and S.G.) independently compiled all clinical and laboratory information from these various sources and compared them with near total matches. Differences were adjudicated based on individual patient chart review and were explained by either missing or updated information from the data warehouse.

Variables. Our data set included three broad classes of variables: 1) demographic variables (age, sex, race, ethnicity and smoking status); 2) clinical variables for each day of hospital encounter (BMI, heart rate, temperature, respiratory rate, O $_2$ saturation, systolic blood pressure, diastolic blood pressure, admission status, discharge status and deaths); and 3) comorbid conditions (CKD, asthma, COPD, hypertension, obesity, diabetes, HIV, sleep apnea and cancer). All three categories were obtained from the patients' electronic medical records, with comorbid conditions defined as an active International Classification of Diseases (ICD)-10 code and vital signs recorded for each patient's given encounter. Although ICD-10 codes represent an international system of categorical variables that are consistent between both practitioners and healthcare systems, we acknowledge that the capture of these data from observational or retrospective cohorts might be less reliable than prospective data collection focusing on specific data elements.

Determining COVID-19 disease severity. A severity scale for COVID-19 was devised by pulmonologists at Mount Sinai based on literature²⁷ and clinical

practice, which defined categories as follows: 1) mild/moderate COVID-19, based on normal/abnormal ($<94\%$) O $_2$ saturation, respectively, or pneumonia on imaging; 2) severe COVID-19, based on use of high-flow nasal cannula (HFNC), non-rebreather mask (NRB), bilevel positive airway pressure (BIPAP) or mechanical ventilation and no vasopressor use, and based on CrCl greater than 30 and alanine aminotransferase (ALT) less than 5 \times the upper limit of normal; and 3) severe COVID-19 with end organ damage, based on use of HFNC, NRB, BIPAP or mechanical ventilation with use of vasopressors, or based on CrCl less than 30, new renal replacement therapy (hemodialysis/continuous veno-venous hemofiltration) or ALT more than 5 \times the upper limit of normal. Clinical notes and imaging reports were reviewed in an effort to establish the patients' COVID-19 disease severity over time. Using a bag-of-words approach to vectorize both clinical notes and image reports, vectors were derived from chest X-ray imaging reports, to reflect the presence of viral pneumonia and worsening respiratory symptoms. Intubation status and O $_2$ therapy modality were obtained by examining the patient clinical notes. The use of endotracheal tube, BIPAP, continuous positive air pressure, HFNC, mechanical ventilator and/or supplemental O $_2$ greater than FiO $_2$ 70% was associated with severe COVID-19. End organ damage was defined by an ALT level greater than 5 \times the upper limit of normal, CrCl less than 30, use of vasopressors and/or new renal replacement therapy.

Alternatively, because of the prevalence in the literature, we also calculated the SOFA score (<https://www.mdcalc.com/sequential-organ-failure-assessment-sofa-score>) for each visit, even if it normally best applied to those patients within the intensive care unit. SOFA score was significantly correlated with our hospital-based severity score ($n=1,450$ pairs, Spearman's $r=0.43$, $P<0.0001$).

Statistical analysis. Patient characteristics were summarized using the standard descriptive statistics: median/IQR for continuous variables and count/percent for categorical variables. Distributions of the cytokine values were assessed and log $_2$ transformed to render the parametric statistical analyses. The cytokines were then categorized at the level of 70 pg ml $^{-1}$, 50 pg ml $^{-1}$, 35 pg ml $^{-1}$, 0.5 pg ml $^{-1}$, 100 mg L $^{-1}$, 1,000 μ g L $^{-1}$ and 1 mg L $^{-1}$ for IL-6, IL-8, TNF- α , IL-1 β , CRP, ferritin and D-dimer, respectively. These stringent cutoffs for cytokines were decided based on empiric testing of various cutoffs within COVID-19 samples, and choosing those rounded above the median of COVID-19 distribution, except for TNF- α where we chose a cutoff based on the upper 99th percentile value of controls because of greater overlap in detection range, whereas those for elevated inflammatory markers were based on 2–3 \times the upper limit of normal detection. The univariate analyses assessed the association of the cytokines and laboratory tests with patient characteristics using the Mann–Whitney U test, the Kruskal–Wallis test and Spearman's rank correlation test as appropriate. Additionally, Deming regressions and Spearman's correlation coefficients were calculated for correlations between the cytokines and laboratory tests. We used multivariable linear regression models to test the association of the cytokine values with patient demographics and comorbidities²⁸. Kaplan–Meier plots along with log-rank tests were conducted to assess the differences in survival probabilities between the high and low levels of each cytokine across the follow-up timeframe, which was calculated from the date of cytokine testing to date of death, discharge or end of follow-up period as appropriate²⁹. The Cox proportional hazards model was used to estimate the hazard of death adjusting for the covariates (for example, patient demographics, comorbidities and laboratory test results), which were determined by the backward elimination method^{30,31}. We assessed the survival model, censoring patients discharged alive and in hospitals. The competing risk model, in which death was the event of interest, live discharge was the competing event and inpatients were censored^{32,33}, was also fitted as a sensitivity analysis. Point estimates (HRs), along with the corresponding 95% CIs, predicted survival probabilities and cumulative incidence curves, were provided. The analyses were performed using two-sided tests and the GraphPad Prism 8.4.2., SAS 9.4 and R 3.6.3 programs.

Validation approaches. We performed internal validation using two methods: 1) ten-fold CV and 2) bootstrap validation (10,000 resamples). The discriminative ability of the Cox proportional hazards regression model was assessed using Harrell's concordance index, denoting the probability that a randomly selected patient with a higher survival time has a higher probability of survival predicted compared to a randomly selected patient with a lower survival time. The goal is to test the model's ability to predict new data, to flag problems like overfitting and selection bias and to give an insight on how the model will generalize to an independent external data set.

In addition, we used an independent validation cohort ($n=231$) and applied the parameter estimates from the primary model (using demographic + comorbidity data as confounders) and the secondary model (using demographic + comorbidity + lab data as confounders) to compute the prognostic index (PI), the linear predictor derived from the model fitted to the derivation cohort. Performance of this external validation model was captured by the models' discrimination and calibration capabilities³⁴. Discrimination was measured by the AUCs over the follow-up time, which were estimated by the statistic c (Harrell's concordance index). Point estimates of 95% CIs and integrated AUCs were computed. Calibration was assessed by plotting Kaplan–Meier curves using the actual survival probabilities in the validation cohort and by comparing them

with the corresponding predicted survival probabilities. Closeness of these two curves is a sign of good calibration. The distribution of the PIs in the original and validation cohorts were presented as histograms and summarized in Extended Data Fig. 4. The similar spread of these distribution provides evidence toward the appropriateness of the validation cohort.

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

Data availability

The data supporting this publication have been made available at ImmPort (<https://www.immport.org>) under study accession [SDY1662](https://www.immport.org). The data set has been de-identified in compliance with the Health Insurance Portability and Accountability Act. ImmPort is a data sharing and data analysis portal for the immunology research community funded by the National Institute of Allergy and Infectious Diseases and the Division of Allergy, Immunology and Transplantation. For further details, refer to the ImmPort user agreement (<https://www.immport.org/agreement>).

Code availability

Scripts used to query the Clarity and Caboodle databases, as well as the statistical analysis, have been uploaded to a GitHub repository: <https://github.com/delvad03/COVID19ELLA>.

References

27. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C. & Di Napoli, R. Features, evaluation and treatment coronavirus (COVID-19). In *StatPearls* (StatPearls Publishing, 2020).
28. Harrell, F. *Regression Modeling Strategies* (Springer, 2001).
29. Rich, J. T. et al. A practical guide to understanding Kaplan–Meier curves. *Otolaryngol. Head. Neck Surg.* **143**, 331–336 (2010).
30. Cox, D. *Analysis of Survival Data* (Chapman & Hall, 1984).
31. Smith, T. & Smith, B. Survival analysis and the application of Cox's proportional hazards modeling using SAS. In *SAS Conference Proceedings* (SAS Global Forum, 2001).
32. Satagopan, J. M. et al. A note on competing risks in survival data analysis. *Br. J. Cancer* **91**, 1229–1235 (2004).
33. Berger, M., Schmid, M., Welchowski, T., Schmitz-Valckenberg, S. & Beyersmann, J. Subdistribution hazard models for competing risks in discrete time. *Biostatistics* **21**, 449–466 (2018).
34. Royston, P. & Altman, D. G. External validation of a Cox prognostic model: principles and methods. *BMC Med. Res. Methodol.* **13**, 33 (2013).

Acknowledgements

Authors thank N. Fernandez for help with the clustergrammer tool and members of the Human Immune Monitoring Center for help with specimen handling. The authors wish to acknowledge R. Pande and M. Putnam at Bio-Techne for helping to provide

instruments and assay kits for ELISA testing in a Clinical Laboratory Improvement Amendments environment in the timeliest possible way during the health crisis; and R. Hyland from Gilead for giving permission to use the remdesivir-related data. S.G., D.M.D.V., S.K.-S., Ma.M., H.-H.H., P.K., A.R. and Mi.M. were supported by National Cancer Institute U24 grant CA224319. S.G. is additionally supported by grants U01 DK124165 and P01 CA190174. T.H.S. was supported by National Institutes of Health K08AI120806. A.R. was supported by U19 AI118610 and U24 AI1186441. Mi.M. was supported by the Fast Grant fund. The Human Immune Monitoring Center and the Institute for Healthcare Delivery Science received support from Cancer Center P30 grant CA196521. S.P. was supported by grant R01 CA244899.

Author contributions

D.M.D.V. and S.K.-S. contributed equally. H.-H.H. and N.D.B. contributed equally. D.M.D.V., S.K.-S., A.R., J.A.A., A.F.-B., D.R.M., J.J., D.R., C.C.-C., S.P., Mi.M. and S.G. contributed to study concept and design. D.M.D.V., S.K.-S., H.-H.H., N.D.B., S.N., B.W., Y.L., T.H.S., D.M., A.S., T.U.M., O.V.O., A.R., P.K., J.A.A., E.S., S.J., Ma.M., A.W.C., A.F.-B., D.R.M., J.J., D.R., K.S., M.F., C.C.-C., S.P., Mi.M. and S.G. contributed to literature search, writing the manuscript and data interpretation. D.M.D.V., S.K.-S., H.-H.H., N.D.B., S.N., B.W., Y.L., T.H.S., D.M., A.S., T.U.M., X.H., M.P., K.T., O.V.O., A.R., P.K., Ma.M., A.W.C., A.F.-B., D.R.M., J.J., D.R., K.S., C.C.-C., S.P., Mi.M. and S.G. participated in data collection and data analysis. D.M.D.V., S.K.-S., H.-H.H., N.D.B., Ma.M., K.S. and S.G. made figures and tables.

Competing interests

S.G. reports consultancy and/or advisory roles for Merck, Neon Therapeutics and OncoMed and research funding from Bristol-Myers Squibb, Genentech, Immune Design, Agenus, Janssen R&D, Pfizer, Takeda and Regeneron. S.P. reports consulting fees from Foundation Medicine and research funding from Celgene and Karyopharm. D.M. reports consultancy and/or advisory roles from Janssen, Celgene, Bristol-Myers Squibb, Takeda, Legend, GlaxoSmithKline, Kinevant and Foundation Medicine. B.W. reports a consulting role at Sanofi Genzyme. J.A.A. reports grants and personal fees from Gilead, grants and personal fees from Merck, grants and personal fees from Janssen, personal fees from Theratech, personal fees from Medicare, grants from Regeneron and grants and personal fees from ViiV, all outside of the submitted work. S.J. reports consulting and/or advisory roles from Janssen, Celgene, Bristol-Myers Squibb, Takeda, Legend and GlaxoSmithKline.

Additional information

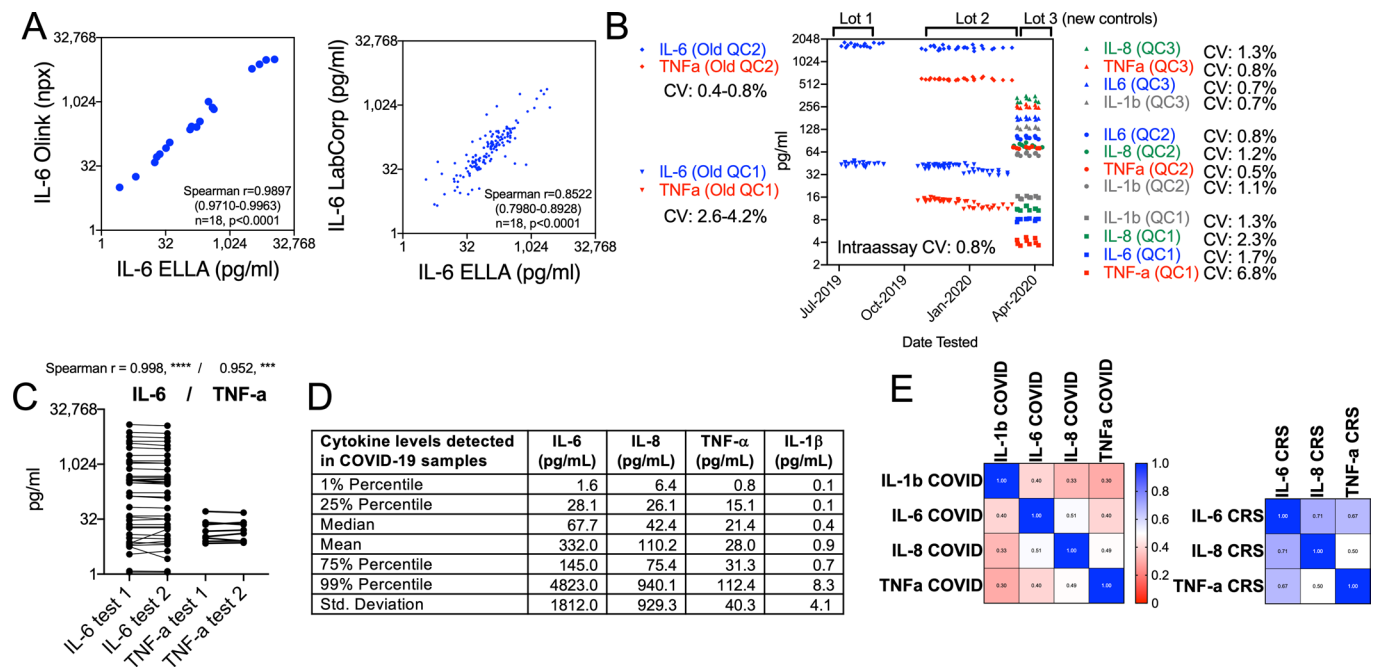
Extended data is available for this paper at <https://doi.org/10.1038/s41591-020-1051-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-020-1051-9>.

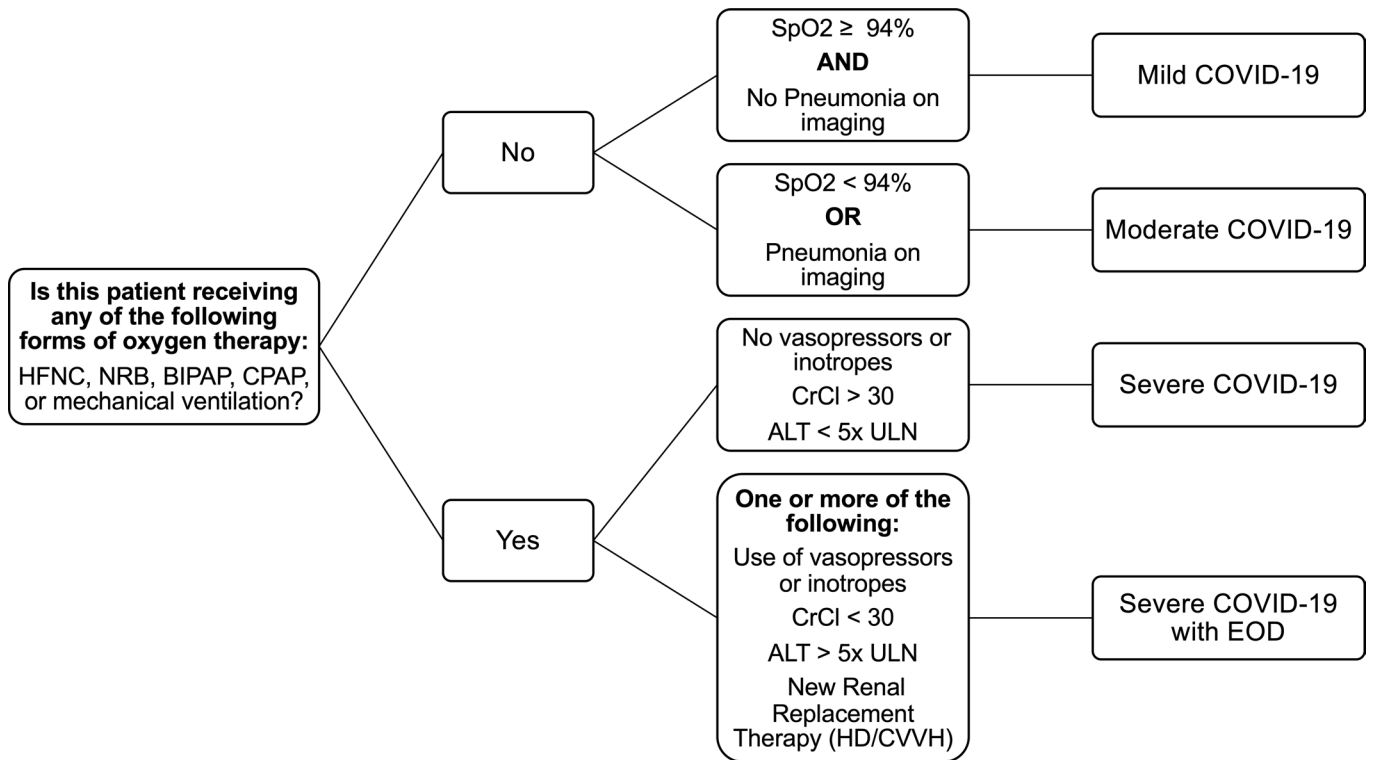
Correspondence and requests for materials should be addressed to S.G.

Peer review information Saheli Sadanand was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

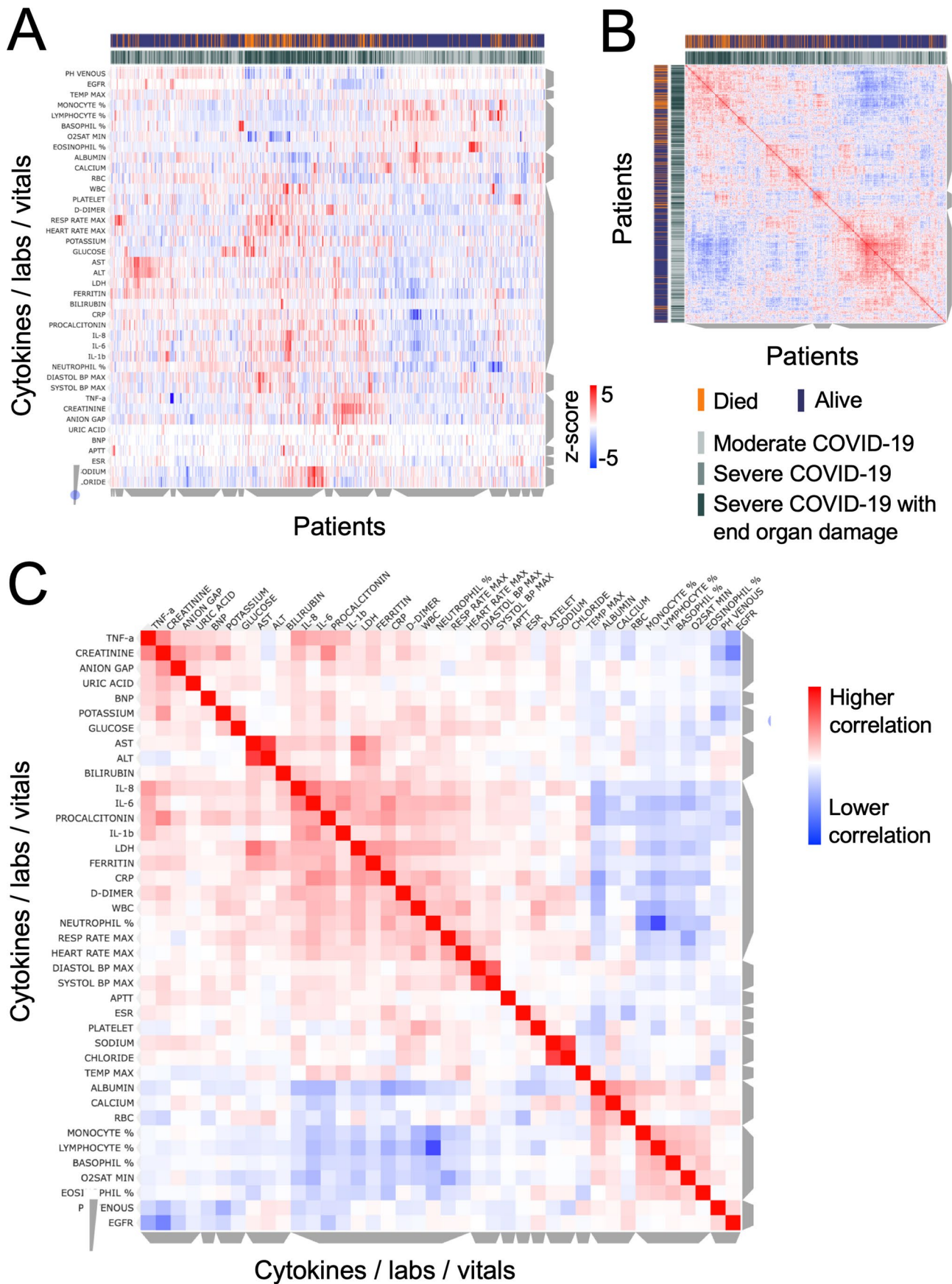
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Sensitivity, specificity, and reproducibility testing of the ELLA platform and cytokine levels and correlation observed in COVID-19 specimens. **a**, Spearman r correlation between IL-6 tested by two other platforms for soluble analyte detection, Olink ($n=18$) and LabCorp ($n=142$), using plasma or serum specimens from CAR-T CRS or COVID-19. **b**, Interassay and intraassay coefficient of variation (CV) of replicates for two recombinant controls used at high or low concentration in each assay for IL-6 and TNF- α using a first set of controls, and for the ELLA panel used in this study using Randox recombinant antigens at three dilution levels over 28 dates tested. **c**, Reproducibility testing including replicates of the same biological specimen from CAR-T samples, with Spearman r indicated for 40 paired samples for IL-6 and 8 paired samples for TNF- α . **d**, Distribution of each cytokine in all COVID-19 samples tested as shown in Fig. 1. **e**, Correlation matrix of IL-1 β , IL-6, IL-8, and TNF- β levels in COVID-19 plasma specimens ($n=1,949$) and in E. multiple myeloma specimens during immunotherapy-related CRS ($n=121$). Scale indicates value of Spearman r correlation. Cytokines levels are less coordinated in COVID-19 than in CAR-T CRS.

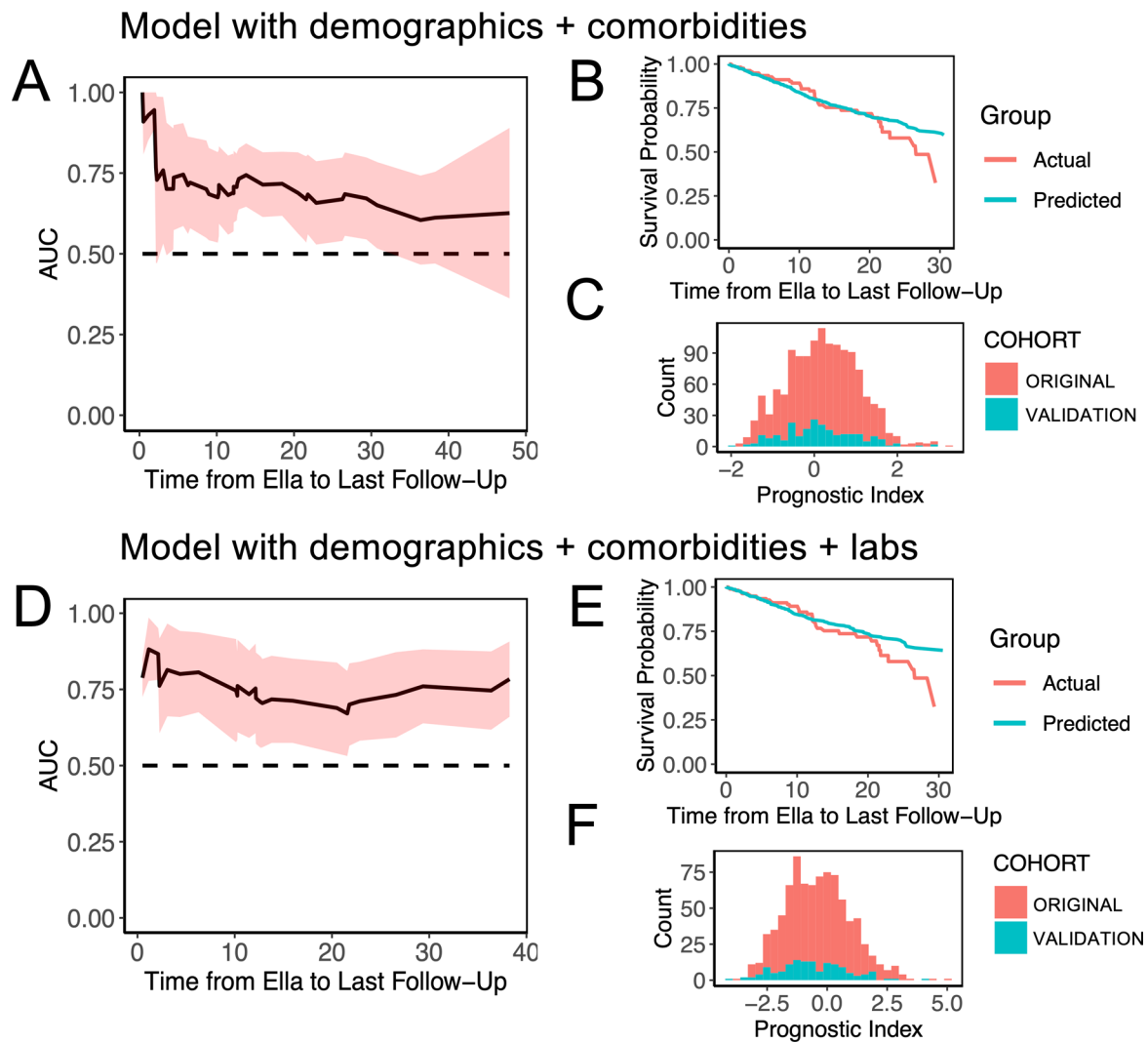


Extended Data Fig. 2 | Flow chart to determine severity. HFNC: high flow nasal cannula; NRB: non-rebreather mask; BiPAP: bilevel positive airway pressure; CPAP: continuous positive airway pressure; SpO₂: oxygen saturation; CrCl: creatinine clearance. ALT: alanine aminotransferase; ULN: upper level of normal; HD/CVVH: hemodialysis / continuous venovenous hemofiltration; EOD: end organ damage.



Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Correlations of clinical and laboratory measurements in individual patients. a, Unsupervised clustering of laboratory measurements in a subset of 1,069 patients with sufficient available information. On the y axis are vitals and laboratory values after z-scoring, and on the x axis are individual patients, using metrics measured from the time point corresponding to the first ELLA cytokine test. Grey bars on the side of the plot indicate clusters of patients or analytes, where cytokines co-cluster with known severity metrics, such as LDH, CRP, ferritin, D-dimer, but also high neutrophil, platelet and white blood counts. Annotations show patients who died in orange, and maximum severity score achieved in gray shades. **b,** Similarity matrix of patients based on analytes and measurements, showing two major clusters, with enrichment in patients who died and had more severe COVID-19 on the upper left. **c,** Similarity matrix of cytokines, lab measurements and vitals, showing IL-6, IL-8, and IL-1 β co-clustering with known inflammatory markers such as LDH, CRP, ferritin, and D-dimer, while TNF- α co-clusters with organ damage markers.



Extended Data Fig. 4 | Validation of the models in an independent cohort. Performance of the model including demographics and comorbidities (**a-c**) or additionally including laboratory metrics (**d-f**) in a validation cohort of 231 patients. **a** and **d**, Pointwise time-dependent area under the curve (AUC) along with the 95% confidence interval were computed to assess the discrimination of the model from cytokine test to last follow-up, with values well above 0.5. **b** and **e**, Calibration was assessed by plotting Kaplan-Meier curves using the actual survival probabilities in the validation cohort and by comparing it with the corresponding predicted survival probabilities. Closeness of these two curves is a sign of good calibration. **c** and **f**, The distribution of the prognostic index in the original cohort (red) and the validation cohort (green) shown as histogram displays a similar spread, providing evidence towards the appropriateness of the validation cohort. Median (IQR) for primary model without labs: original cohort: 0.2401 (-0.3874, 0.8168), validation cohort: 0.0697 (-0.4785, 0.7722). Median (IQR) for model including labs: original cohort: -0.3671 (-1.3246, 0.5021), validation cohort: -0.6642 (-1.5348, 0.3444).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data was processed using R version 3.6.1 (Vienna Austria). Large tables were read-in and written using R packages tidyverse (v 1.3.0), reshape2 (v 1.4.4), and readxl (v 1.3.1). Clinical data was extracted from the Epic electronic health record (Verona, WI) using Epic Hyperspace (August 2019), Epic Clarity (February 2020) and Epic Caboodle (February 2020) databases via connecting to Oracle (18c Enterprise Edition Release 18.0.0.0.0) and SQL server (Microsoft SQL Server 2016 (SP2-CU11) (KB4527378) - 13.0.5598.27 (X64)) databases respectively.

Data analysis

SAS 9.4 (PROC GENMOD, PROC LIFETEST and PROC PHREG), GraphPad Prism 8.4.2, and R 3.6.3. (Package: ggplot2)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data supporting this publication is available at ImmPort (<https://www.immport.org>) under study accession SDY1662.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|-----------------|--|
| Sample size | A two-sided log-rank test with an overall sample size of 674 patients (337 in each group) achieves 80% power at a 0.05 type I error to detect a hazard ratio of 2.29 when the proportion of death in the low group is 20%. In this study, we had more than 1,200 patients so that we can have enough power to detect the potential slight smaller effect sizes after adjusting for the covariates. Please refer to response to Reviewer 2 comments. We queried all available health records. |
| Data exclusions | The patients with the cytokines measured after the date of last follow-up and or missing status (discharge, death, or hospitalization) were excluded. 1,298 patients had cytokines measured on or before the date of last follow-up. Among them, 1,245 patients had the measurements of all four cytokines. Exclusion criteria were pre-established. Our aim was to look at patients with both an indication of SARS-CoV-2 infection and ELLA cytokine data availability. |
| Replication | This observational study is of a hospitalized cohort during the COVID-19 pandemic in New York City. We validated the initial findings in a second cohort. -- The ELLA platform has 3 built in replicates. |
| Randomization | This is not a clinical trial. Established clinical laboratory test data was used. |
| Blinding | This is not a clinical trial. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

| n/a | Involved in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Human research participants

Policy information about [studies involving human research participants](#)

| | |
|----------------------------|--|
| Population characteristics | Please see Table 1. Patient characteristics for details of human subjects research participants. |
| Recruitment | Subjects were not recruited into this study, therefore, self-selection bias does not apply. Data was collected as part of normal clinical care. ELLA was ordered as a clinical test as part of standard clinical care at the MSHS during the SARS-CoV-2 epidemic. Clinical tests were ordered by the treated physician. |
| Ethics oversight | This research was reviewed and approved by the Human Research Protection Program at the Icahn School of Medicine at Mount Sinai (ISMMS), a comprehensive system to ensure the protection of the rights and welfare of subjects in human research. The Program for the Protection of Human Subjects (PPHS) is a key component of ISMMS' efforts to ensure human subject protections. It supports our researchers in assuring the ethical conduct of research and compliance with federal, state, and institutional regulations, and provides a professional office staff to assist both investigators, participants, and five Institutional Review Boards (IRBs). |

Note that full information on the approval of the study protocol must also be provided in the manuscript.