



OPEN Predicting severity in COVID-19 disease using sepsis blood gene expression signatures

Arjun Baghela¹, Andy An¹, Peter Zhang², Erica Acton³, Jeff Gauthier⁴, Elsa Brunet-Ratnasingham^{5,6}, Travis Blimkie¹, Gabriela Cohen Freue⁷, Daniel Kaufmann^{6,8}, Amy H. Y. Lee³, Roger C. Levesque⁴ & Robert E. W. Hancock^{1,2}✉

Severely-afflicted COVID-19 patients can exhibit disease manifestations representative of sepsis, including acute respiratory distress syndrome and multiple organ failure. We hypothesized that diagnostic tools used in managing all-cause sepsis, such as clinical criteria, biomarkers, and gene expression signatures, should extend to COVID-19 patients. Here we analyzed the whole blood transcriptome of 124 early (1–5 days post-hospital admission) and late (6–20 days post-admission) sampled patients with confirmed COVID-19 infections from hospitals in Quebec, Canada. Mechanisms associated with COVID-19 severity were identified between severity groups (ranging from mild disease to the requirement for mechanical ventilation and mortality), and established sepsis signatures were assessed for dysregulation. Specifically, gene expression signatures representing pathophysiological events, namely cellular reprogramming, organ dysfunction, and mortality, were significantly enriched and predictive of severity and lethality in COVID-19 patients. Mechanistic endotypes reflective of distinct sepsis aetiologies and therapeutic opportunities were also identified in subsets of patients, enabling prediction of potentially-effective repurposed drugs. The expression of sepsis gene expression signatures in severely-afflicted COVID-19 patients indicates that these patients should be classified as having severe sepsis. Accordingly, in severe COVID-19 patients, these signatures should be strongly considered for the mechanistic characterization, diagnosis, and guidance of treatment using repurposed drugs.

In the ongoing COVID-19 pandemic (2020–2022), it is increasingly evident that many of the millions of individuals who die with severe disease succumb due to sepsis, a life-threatening dysfunctional response to infection. Studies show that many or most individuals who died from severe COVID-19 infections had acute respiratory distress syndrome (ARDS) and high Sequential Organ Failure Assessment (SOFA) scores (ranging from 2.6 to 12 among patients with life-threatening disease)^{1,2}. Furthermore, more extensive phenotyping studies have suggested that features of sepsis such as considerable organ damage (e.g., lung and kidney injury) and long-term depressed mental health and immune responses are prevalent in severe COVID-19 patients^{3,4}. The possibility has been discussed that this is viral-driven sepsis⁵, but the involvement of bacterial infections in at least 50% of severely-afflicted COVID-19 patients has also been established^{6,7}.

The characterization of COVID-19 disease as a type of sepsis has important clinical implications for the prognosis and treatment of severely-afflicted COVID-19 patients⁸. The clinical criteria (e.g., SOFA, Multiple Organ Dysfunction Score [MODS]), biomarkers, and immunomodulatory therapies proposed for the management of sepsis can likely be generalized to COVID-19 if the severe manifestation of this disease is a type of sepsis. Unfortunately, sepsis itself has proven difficult to define using clinical criteria, since symptomatology ranges from non-specific/mild presentation to organ dysfunction, the requirement of mechanical ventilation and life support, and death⁹. Whole blood transcriptomics has proven useful in characterizing dysregulated host immune responses during sepsis and, importantly, has provided novel opportunities to define the sepsis syndrome using

¹Department of Microbiology and Immunology, University of British Columbia (UBC), Vancouver, Canada. ²Asep Medical, Vancouver, Canada. ³Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, Canada. ⁴Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Microbiologie-Infectiologie et d'immunologie, Université Laval, Quebec, QC, Canada. ⁵Département de Microbiologie, Infectiologie Et Immunologie, Université de Montréal, Montreal, Canada. ⁶Centre de Recherche du CHUM, Montreal, QC, Canada. ⁷Department of Statistics, UBC, Vancouver, Canada. ⁸Département de Médecine, Université de Montréal, Montreal, Canada. ✉email: bob@hancocklab.com

multi-molecule signatures¹⁰. These signatures have enabled the development of more specific criteria to identify, diagnose, and categorize sepsis patients. For example, we recently characterized, in emergency room (ER) patients, a set of gene expression signatures reflective of sepsis endotypes (a recent paradigm to separate patients into coherent and mechanistically distinct subgroups) as well as endotype-independent global markers of severity predictive of impending organ dysfunction and mortality^{11,12}.

Many studies have used transcriptomics to characterize the specific gene expression events dysregulated in severe COVID-19 patients, like those performed in all-cause sepsis contexts. These have suggested that COVID-19 specific signatures/mechanisms, in part reflecting innate immune responses and canonical hallmarks of inflammation in neutrophils, monocytes, and T cells, can likely be exploited for diagnostic and prognostic value in COVID-19 disease¹³. However, far fewer studies have explored the gene expression profiles of severe COVID-19 patients in the context of known molecules dysregulated in all-cause (but frequently bacterial-infection triggered) sepsis. This would provide an explicit link to the underlying biological programs defining the sepsis syndrome. In this study, we recruited patients with confirmed COVID-19 infection and a broad range of disease severity from various centers in Quebec, Canada to determine if COVID-19 patients displayed a significant dysregulation of established signatures. We show that all-cause sepsis signatures, specifically those reflecting cellular reprogramming (CR), organ dysfunction, and mortality, are mediators of COVID-19 pathogenesis and discriminate between COVID-19 severity groups. Furthermore, COVID-19 patients fit into the same mechanistic groupings/endotypes that are evident in all-cause sepsis. These results suggest that mechanisms underlying the severity of COVID-19 are those found in sepsis, and can be represented by the specific gene expression signatures that define this syndrome.

Methods

Study design and patient recruitment. We sought to characterize global transcriptional changes that occur through a range of severities from moderate to severe COVID-19 disease (as defined by requirement for respiratory support) and to determine whether these changes overlapped with those observed during all-cause sepsis. The Biobanque Québécoise de la COVID-19 (BQC-19; Quebec COVID-19 Biobank; <https://en.quebecovidiobank.ca/>) recruited, in 10 hospitals, adult and pediatric patients (across ERs, wards, ICU, and outpatient clinics) with confirmed SARS-CoV-2 infection in a province-wide initiative to collect, store, and share samples and data related to the Covid-19 pandemic¹⁴. The study was approved by the respective institutional review boards and written informed consent was obtained from all participants or, when incapacitated, their legal guardian before enrollment and sample collection.

COVID-19 severity was assessed by the extent of respiratory support received at sampling, with patients assigned to groups comprising severe (evidence of organ failure and required ECMO, mechanical ventilation, or non-invasive ventilation), intermediate (needed supplemental oxygen through a nasal cannula), or moderate (required no supplemental oxygen and had milder symptoms), as well as recording eventual mortality. To enable retrospective association with gene expression data, various clinical and demographic metadata were collected, including age, sex, duration of hospital stay, non-COVID active infections, etc. (Table 1). The Clinical Research Ethics Board of the University of British Columbia (UBC; approval number H17-01208) and Comité d'éthique

Parameter	Quebec hospital-wide (wards and ICU) cohort (N = 124)
Age	61.3 ± 1.35 (124)
Sex, female	38.7% (48/124)
Duration of illness before hospital arrival (days)	6.6 ± 0.37 (118)
Hospital stay duration (days)	23.2 ± 2.0 (124)
On Antibiotics	80.0% (96/120)
Non-COVID active infection	35.8% (44/123)
Sampling group	Early 61.3% (76 patients); late 38.7% (48 patients)
Existing lung disease	10.5% (13/124)
Existing kidney disease	20.2% (25/124)
Existing heart disease	11.2% (13/124)
Existing liver disease	0.81% (1/124)
Respiratory support at sampling	Severe, 43.5% (54/124); intermediate, 22.6% (28/124); moderate, 33.9% (42/124)
SOFA scores	8.02 ± 0.54 (54)
ICU admission	44.4% (55/124)
Mortality (in hospital)	16.1% (20/124)

Table 1. Sepsis severity and outcomes of the cohorts with confirmed COVID-19. The mean value ± standard error is presented for numerical variables (total observations in brackets). Categorical variables are presented as % positive observations (relative to total observations). Early samples were taken 1–5 days post admission while late samples 6–20 days post admission. Numbers in brackets are the numbers of patients in each category. Respiratory support definition—severe: required extracorporeal membrane oxygenation (ECMO), Intubation, or Non-invasive ventilation; intermediate: required supplemental oxygen through nasal cannula; moderate: no supplemental oxygen required.

de la recherche du Centre hospitalier de l'Université de Montréal (CUM-CHUM; approval numbers MP-02-2020-8929 and 19.389) provided ethics approval for all sequencing and bioinformatics studies, carried out in a manner blinded to patient identity. All methods described were performed in accordance with the relevant guidelines and regulations.

Blood collection and transcriptome sequencing (RNA-Seq). During the usual blood sample collections in hospital patients, an additional 2.5 ml of blood was collected for RNA-Seq. Blood was collected into PAXgene Blood RNA tubes (BD Biosciences; San Jose, CA, USA) to ensure stabilization of intracellular RNA¹⁵. Total RNA was extracted using the PAXgene Blood RNA Kit (Qiagen; Germantown, MD, USA). Quantification and quality measures for total RNA were obtained using a LabChip GXII instrument (Waltham, MA, USA). Poly-adenylated RNA was captured using NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB; Ipswich, USA) and cDNA libraries were prepared using the NEBNext RNA First Strand Synthesis Module, NEBNext Ultra Directional RNA Second Strand Synthesis Module, and NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB; Ipswich, USA). RNA-Seq was then performed at a depth of 50 M reads/sample on an Illumina NovaSeq 6000 S4 instrument of 100 base-pair long paired-end sequence reads (excluding adapter/index sequences).

A standard data processing protocol was used, including quality control using fastqc (v0.11.9)¹⁶ and multiqc (v1.6)¹⁷, alignment to the human genome (Ensembl GRCh38.104) using STAR (v2.7.9a)¹⁸, and read count assessments using htseq-count (v0.11.3)¹⁹. Finally, globin genes and genes with fewer than 10 counts were removed from count tables, and we ensured that analyzed samples had more than one million total read counts. Raw sequencing data are available at NCBI GEO (Accession Number GSE221234). When performing an analysis other than differential gene expression analysis (which typically requires raw counts as input), the variance stabilizing transformation (VST) from the DESeq2 (v1.22.2) R package was performed to render counts homoscedastic and normalized for varying library sizes²⁰. Following transformation, technical variation due to sequencing batch was removed using ComBat within the Surrogate Variable Analysis R package (v3.30.1)²¹.

Bioinformatics and statistical analysis. Mechanisms of COVID-19 severity were determined using differential expression analysis by comparing severity groups based on the respiratory support required and impending mortality. Differential expression analysis was performed using DESeq2, with genes considered differentially expressed (DE) if they displayed $\geq \pm 1.5$ -fold change and Benjamini–Hochberg adjusted p value ≤ 0.05 (Wald test). Functional characterization of DE genes was performed using an over-representation/enrichment analysis of Reactome pathways²² and MSigDB hallmarks²³. A pathway/hallmark was considered enriched if it displayed an adjusted p value ≤ 0.05 (hypergeometric test).

We explored whether these patients demonstrated dysregulation of previously-described all-cause sepsis severity signatures^{12,24}. These signatures were identified in a geographically diverse cohort of early all-cause sepsis patients and further validated in two independent cohorts of ER and ICU patients (including non-Covid-19/bacterial and Covid-19 sepsis patients). This allowed us to validate the use of all-cause sepsis signatures and further examine whether severe COVID-19 is a type of sepsis. The supervised DE gene-set algorithm Rotation Gene Set Test (ROAST), provided within the limma R package (v3.38.3)^{25,26}, was used to assess whether these signatures were significantly dysregulated between COVID-19 severity groups.

Sepsis endotypes were also examined within the data¹². Endotypes are subtypes of a disease defined by unique pathophysiology and underlying mechanisms. Predicting endotype status in the current cohort involved calculating a sample-wise gene-set enrichment score for each independent 200-gene endotype signature. Specifically, the over- or under-expression of signatures in individual patients was estimated using the GSEA R package (v1.30.0); this assessed expression differences of a gene set when compared to all other genes (i.e., genes not in the set) within each sample/patient²⁷. Then patients were classified to the endotype exhibiting the highest expression or enrichment of the respective endotype signature. Clinical parameters were statistically compared between predicted endotypes using Kruskal–Wallis and Chi-Squared tests to compare numerical and categorical variables, respectively.

The funders (Canadian Institutes for Health Research; FDN-154287) and the Fonds de recherche du Québec (FRQ) COVID-19 Biobank had no role in the manuscript's analysis, interpretation, or writing.

Results

Mechanisms of sepsis severity and mortality in COVID-19 patients. To identify COVID-19 specific severity mechanisms, we initially compared the whole blood gene expression profiles associated with defined severity groups from a cohort of 124 patients recruited at various times relative to hospital admission. Patient severity was assessed using two measures, one based on the requirement for respiratory support and the second based on whether a patient died in hospital. Based on principal component analysis (PCA) of global gene expression profiles, patients could be well separated according to respiratory support requirement group and eventual in-hospital mortality on two dimensions (Supplemental Fig. 1a). The association of several parameters (including respiratory support requirement group, various comorbidities, and time of sampling relative to hospital admission) with gene expression principal components was also explored using multiple linear regression (Supplemental Fig. 1b). These results indicated strong association with the second principal component, PC2, for eventual mortality (p value = 0.000059; multiple linear regression) and respiratory support group (p value = 0.0062), with PC2 explaining 17.8% of the total gene expression variance. Interestingly, other influences on gene expression, including sex, non-COVID-19 active infections, active cancer diagnosis, etc., were not associated with the most variable PCs (i.e., PC1–4).

Differential expression analysis comparing patient severity groups, as defined by need for respiratory support, indicated several thousands of genes were dysregulated in severe COVID-19 disease. Most notably, when

comparing severe vs. moderate patients, there were 2671 differentially expressed (DE) genes, with 1794 up-regulated and 877 down-regulated ($\geq \pm 1.5$ -fold change; adjusted $p \leq 0.05$). These were analyzed for over-representation of Reactome²² and Hallmark gene-sets²³, which capture pathways and well-defined biological processes, respectively. Up-regulated genes differentiating severe vs. moderately ill patients were enriched in several pathways, including those involving neutrophil degranulation, interferon $\alpha/\beta/\gamma$ signaling, interleukin (IL)-4 and IL-13 signaling, oxidative stress, and signaling by NOTCH. Furthermore, canonical inflammation hallmarks such as TNF- α signaling via NF- κ -B and the Inflammatory response were also enriched (Fig. 1). Down-regulated DE genes were enriched in other pathways/hallmarks, including adaptive immune pathways such as programmed cell death-1 (PD-1) signaling, generation of second messenger molecules, TCR signaling, and allograft rejection. These results indicated that while anti-viral responses were enhanced, T-cell responses and activity were likely reduced during severe COVID-19 disease, consistent with severe all-cause sepsis^{12,28}. The severe vs. intermediate comparison also showed a significant dysregulation of gene expression (535 up-regulated, 159 down-regulated); with enriched pathways/hallmarks including neutrophil degranulation, interferon $\alpha/\beta/\gamma$ signalling, coagulation, and complement. The intermediate vs. moderate comparison yielded relatively few DE genes (136 up-regulated and 33 down-regulated) and enriched pathways/hallmarks, indicating these groups were most similar based on gene expression (or conversely that the intermediate group showed some heterogeneity).

When comparing patients who succumbed in the hospital to those who survived, there were 1006 up-regulated and 702 down-regulated genes. The majority of these genes overlapped with the DE genes obtained when comparing severe vs. moderate patients (1122 overlapping genes) (Fig. 1). As such, neutrophil degranulation, interleukin signaling, down-regulation of T cell signaling pathways, and inflammatory hallmarks were commonly



Figure 1. Functional characterization of the gene expression differences between severity groups and eventual mortality in Quebec COVID-19 patients. Shown are the top 10 enriched pathways/hallmarks for each comparison, direction, and gene set database assessed (i.e., lowest p values). Severity groups were assessed using the respiratory support required at time of sampling with 54 severe, 28 intermediate, and 42 moderate patients. There were 20 patients who died and 100 patients who survived. The severe vs. moderate, severe vs. intermediate, and intermediate vs. moderate, died vs. survived, comparisons yielded 2671 (1794 up, 877 down), 694 (535 up, 159 down), 169 (136 up, 33 down), and 1708 (1006 up, 702 down) DE genes, respectively (displaying $\geq \pm 1.5$ -fold change; adjusted $p \leq 0.05$). Age, sex, batch, and time of sampling relative to hospital admission were included in the underlying DESeq2 model to correct for their contribution to gene expression. A post-hoc power analysis (performed with the online tool RnaSeqSampleSize⁵³) using parameters estimated from previous RNA-Seq studies, indicated our study was sufficiently powered ($p = 0.8$) to detect DE gene differences in gene expression (at least 15% of transcriptome) between patient groups.

enriched between respiratory support severity and eventual mortality comparisons. Interestingly, heme metabolism, cholesterol homeostasis, and p53 pathways were uniquely enriched in DE genes from the mortality comparison, indicating that additional mechanisms became prominently dysregulated when mortality was imminent. Indeed, heme metabolism was the most significantly enriched hallmark among genes DE between patients who died vs. survived, and was downregulated (Fig. 1). This suggests that the down-regulation of heme metabolism is likely a specific feature of severe Covid-19 disease and was most evident among patients who succumbed to this disease. We further examined gene expression data of previously published COVID-19 positive and negative patients from Toronto, Ontario, who were recruited on the first day of ICU admission for COVID-19 specific pathways (GEO accession GSE185263)¹². Highly enriched interferon- $\alpha/\beta/\gamma$ signaling (up-regulated) and heme metabolism (down-regulated) was apparent in COVID-19 positive patients when compared to negative patients as well as in COVID-19 positive compared to uninfected healthy controls (Supplemental Fig. 2).

The common enrichment of the neutrophil degranulation pathway across severity comparisons indicated that neutrophil activity and/or abundance was a marked feature of severe COVID-19 sepsis. This feature has been described in previous studies of all-cause sepsis, with several neutrophil and granulocyte genes and pathways being dysregulated in all-cause sepsis patients progressing to organ dysfunction/failure^{12,29}. The ratio of neutrophils to lymphocytes cellular abundance in patients has been touted as a moderately accurate (AUC \approx 0.695) and easily measurable prognostic biomarker of sepsis severity and outcomes in patients with advanced sepsis³⁰. We estimated the neutrophil to lymphocyte ratio using the gene expression-based deconvolution program CIBERSORT³¹ and from cell differential values measured in hospital in order to explore their association to respiratory support group (Supplemental Fig. 3). It was clear that patients with severe and intermediate COVID-19 had the highest neutrophil to lymphocyte ratio which significantly differed from that of moderate patients ($p = 9.2 \times 10^{-7}$ by the Kruskal–Wallis test using CIBERSORT estimated ratios).

Although our initial motivation was to identify COVID-19 specific markers of patient severity, it was apparent from our analyses that existing comorbidities in patients contributed to the observed gene expression values (Supplemental Fig. 1b), including chronic renal, lung and heart insufficiencies and active cancer. It is well-accepted that existing comorbidities influence the risk of sepsis and sepsis severity³², however, few studies have examined the role of comorbidities on the transcriptomic profiles in COVID-19 sepsis. Accordingly, we performed DE analysis, followed by pathway/hallmark analysis, to identify mechanisms dysregulated between COVID-19 patients with existing chronic insufficiencies vs. those without (Supplemental Fig. 4). The largest dysregulation observed was among patients with chronic lung insufficiencies, which is notable considering SARS-CoV-2 infection is characterized by progression of the infection into the lower respiratory tract (i.e., lung tissue). This comparison yielded down-regulated MSigDB hallmarks, including complement, coagulation, and inflammatory responses. Interestingly, patients with chronic renal insufficiencies, cf. those with no chronic renal issues, also exhibited dysregulation of unique mechanistic subsets, including an up-regulation of interferon $\alpha/\beta/\gamma$ signaling, which was also observed among comparisons of Covid-19 severity. When comparing those with chronic heart insufficiencies vs. not, few DE genes and pathways/hallmarks arose. These results indicate that overall specific gene expression differences occurred due to pre-existing comorbidities during COVID-19 disease, highlighting the possibility of adjusting therapeutic interventions according to comorbidity status.

COVID-19 patients displayed dysregulation of all-cause sepsis severity signatures. Several mechanisms were dysregulated between COVID-19 severity groups, and many of these could be biomarkers of COVID-19 pathogenesis and thus valuable for patient prognostication. Some of these mechanisms have been implicated in sepsis, including those reflecting neutrophil activity, inflammatory cytokines, and adaptive immune pathways strengthening the notion that severe COVID-19 is a type of sepsis. To further indicate the extent to which severe COVID-19 overlaps with all-cause sepsis, we took the reverse approach, specifically determining whether prognostic gene expression signatures uncovered in all-cause sepsis cohorts are dysregulated in COVID-19 patients.

We previously described reduced 8–12 gene mechanistic signatures of sepsis severity uncovered in all-cause sepsis training cohorts^{11,12}. These included the cellular reprogramming (CR; 8 genes), Organ Dysfunction (12 genes), and Mortality (10 genes) signatures, which were identified and validated in early sepsis patients at first clinical presentation to the ER as well as in the ICU. The CR and Organ Dysfunction signatures were shown to be highly accurate in predicting impending severe sepsis and organ dysfunction. Similarly, the Mortality signature was strongly associated with 30-day mortality and reasonably predictive of impending mortality. We determined whether these signatures were dysregulated between COVID-19 severity groups using ROAST, a supervised gene-set DE algorithm that assesses whether a set of genes, as a unit, is significantly up- or down-regulated²⁵.

When analyzing all patients among the Quebec cohort, the CR, Organ Dysfunction, and Mortality signatures were strongly associated with the assessed endpoints (Table 2). In particular, when comparing severe vs. moderate patients each signature displayed p values < 0.0001 , indicating a strong degree of signature dysregulation between these patient groups. This suggested that all-cause sepsis signatures were also highly relevant to COVID-19 pathogenesis in severely-afflicted patients. It is important to note that this cohort was comprised of patients sampled early and late in their hospital stays, and, as indicated in Supplemental Fig. 1b, sampling to admission time played a significant role in gene expression variability. Accordingly, we assessed the gene-set level DE of the severity signatures among patients' groups separated by sampling time (i.e., early and late). This was done to minimize the role that disease progression might have had in the performance of the signatures. Patients were stratified into early (sampled within 1–5 days) and late (sampled between 6 and 20 days) post-hospital admission groups to assess whether gene expression signatures exhibited altered behaviour when accounting for this source of heterogeneity. The results indicated that across early and late sampled patients, the signatures were still significantly dysregulated between severe vs. moderate and dead vs. surviving patients. Severe vs. intermediate

Cohort	Comparison	Statistical significance (p value)		
		CR	Organ dysfunction	Mortality
All	Severe (54) vs. moderate (42)	<0.0001	<0.0001	<0.0001
	Severe (54) vs. intermediate (28)	<0.0001	0.0058	0.063
	Intermediate (28) vs. moderate (42)	<0.0001	0.0021	0.0045
	Died (20) vs. survived (104)	0.0014	0.0023	<0.0001
Early sampled (0–5 days)	Severe (28) vs. moderate (29)	<0.0001	0.00040	<0.0001
	Severe (28) vs. intermediate (19)	0.053	0.47	0.037
	Intermediate (19) vs. moderate (29)	0.00020	0.0057	0.025
	Died (11) vs. survived (65)	0.012	0.0095	0.00030
Late sampled (6–20 days)	Severe (26) vs. moderate (13)	<0.0001	<0.0001	0.0014
	Severe (26) vs. intermediate (9)	0.0090	0.0020	0.44
	Intermediate (9) vs. moderate (13)	0.29	0.25	0.076
	Died (9) vs. survived (39)	0.043	0.13	0.0043

Table 2. Gene-set differential expression of the reduced cellular reprogramming/CR (8 genes), organ dysfunction (12 genes), and mortality (10 genes) signatures in COVID-19 patient cohorts. The gene sets were assessed for significant dysregulation between patients using ROAST. Statistically significant gene-set DE is indicated by bold typeface (FDR adjusted $p \leq 0.05$).

and intermediate vs. moderate patients exhibited lesser dysregulation, and this was most evident in late sampled patients. The gene set level dysregulation of the full CR, Organ Dysfunction, and Mortality signatures¹² was also assessed (Supplemental Table 1), and displayed similar results to those obtained using the reduced signatures (Table 2).

We also determined the predictive accuracy of the CR, Organ Dysfunction, and Mortality signatures when used to directly predict patient severity groups. Logistic regression models, for which the reduced signature gene sets provided input into the model as covariates/predictors (Fig. 2a), were trained to predict respiratory support group and eventual mortality. In the case of respiratory support group, a binary classification scheme was trained to predict severe vs. intermediate + moderate patients. We combined the intermediate and moderate patients because their gene expression profiles were most similar when analyzed using differential expression analysis. Due to the modest size of these predictive reduced signatures (8–12 genes), they reflected subsets with the greatest potential for translation, since smaller signatures can be more easily assessed in clinics using qRT-PCR. A tenfold repeated cross validation was used to estimate predictive performance. The Organ Dysfunction (mean AUC = 0.79, sensitivity = 0.71, specificity = 0.68) and CR (mean AUC = 0.86, sensitivity = 0.80, specificity = 0.79) signatures displayed good performance in early sampled patients. Among late sampled patients, the Organ

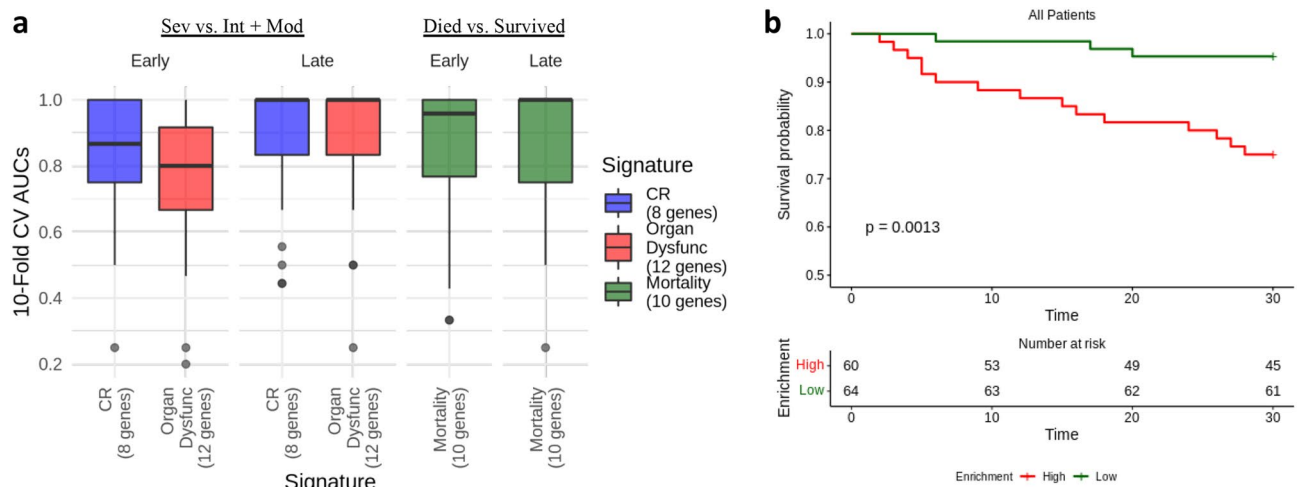


Figure 2. Predictive performance of the reduced CR, Organ Dysfunction, and Mortality signatures among patients of the COVID-19 patient cohorts. **(a)** Accuracy/AUCs across a tenfold repeated cross validation scheme (i.e., total 100 folds) using logistic regression. **(b)** Kaplan Meier curves of patients with High or Low Mortality signature enrichment in all patients (i.e., early and late sampled patients). Time to mortality was assessed over 30 days, with the baseline time being the day of sampling. Patients were stratified into either High (score ≥ 0 ; $n = 60$) or Low (score < 0 ; $n = 64$) signature expression based on GSEA enrichment statistics, and this was used as a predictor of 30-day survival.

Dysfunction (mean AUC = 0.92, sensitivity = 0.80, specificity = 0.77) and CR (mean AUC = 0.91, sensitivity = 0.83, specificity = 0.81) signatures also displayed excellent performance.

The Mortality signature also displayed good performance in predicting eventual mortality in early (mean AUC = 0.86, sensitivity = 0.77, specificity = 0.75) and late (mean AUC = 0.89, sensitivity = 0.78, specificity = 0.86) sampled patients. We also explored the association of the Mortality signature with 30-day mortality using Kaplan Meier survival curves across all patients. The patient-level enrichment/expression scores were estimated for the Mortality signatures using GSVA that summarizes the expression of a gene set when compared to all other genes (i.e., genes not in the set) within each sample. Patients were separated into High (enrichment score ≥ 0 ; $n = 60$) and Low (enrichment score < 0 ; $n = 64$) signature enrichment categories to test their association with 30-day mortality. There was a significant association of high mortality signature enrichment with 30-day mortality (full mortality signature: log-rank p value = 0.00062; reduced mortality signature: log-rank p value = 0.0013) (Fig. 2b).

COVID-19 patients exhibited evidence of all-cause sepsis endotypes. Our group also recently described, in ER patients at first clinical presentation, five sepsis endotypes. Endotypes are distinct disease groupings characterized by unique pathobiological mechanisms and DE gene expression signatures¹². Importantly, endotypes are hypothesized to accurately explain the heterogeneous nature of the sepsis syndrome in contrast to earlier paradigms considering sepsis as a single disease. The early sepsis endotypes included two poor prognosis endotypes, namely the Neutrophilic-Suppressive (NPS) and Inflammatory (INF) endotypes, that were respectively associated with neutrophil activation/immune suppression and increased pro-inflammatory responses. Three additional fair-prognosis endotypes were also uncovered, namely the Innate Host Defences (IHD), Interferon (IFN), and Adaptive (ADA) endotypes. To predict endotype status in other cohorts, five 200-gene endotype-specific signatures were derived. Evidence of sepsis endotypes among COVID-19 patients would further support the notion that severe COVID-19 disease is the same as, or similar to, all-cause sepsis.

Initially, to characterize endotypes in patient whole blood RNA-Seq data, we determined, using a data-driven approach, whether the gene expression data supported the partitioning of these patients into 5 discrete clusters. Cluster stability metrics inferred that 4–6 clusters could be reliably discerned among all patients analyzed together and also among patients separated into early and late sampled groups (Supplemental Fig. 5). We next sought to classify patients into one of the five previously-described all-cause sepsis endotypes. A better assessment of severity signatures was obtained by stratifying patients into the early and late cohort patients, which partially resolved gene expression heterogeneity due to sampling time relative to hospital admission. Accordingly, endotype status was separately predicted in early and late sampled patients (Fig. 3) by calculating a sample-wise gene-set enrichment score for each 200-gene endotype-specific signature using GSVA (v1.30.0). Patients were then assigned to the endotype for which they exhibited the highest expression/enrichment score; this is depicted in Fig. 3 as blocks of red (demonstrating enrichment) in the bottom five rows of each subfigure.

Across both early and late sampled patients, endotype classification worked well to separate patients into five endotypes of similar size. In early sampled patients, these (predicted) endotypes positively correlated with assignment to severe respiratory support requirement group (p value = 0.0015), qSOFA (p value = 0.045), and

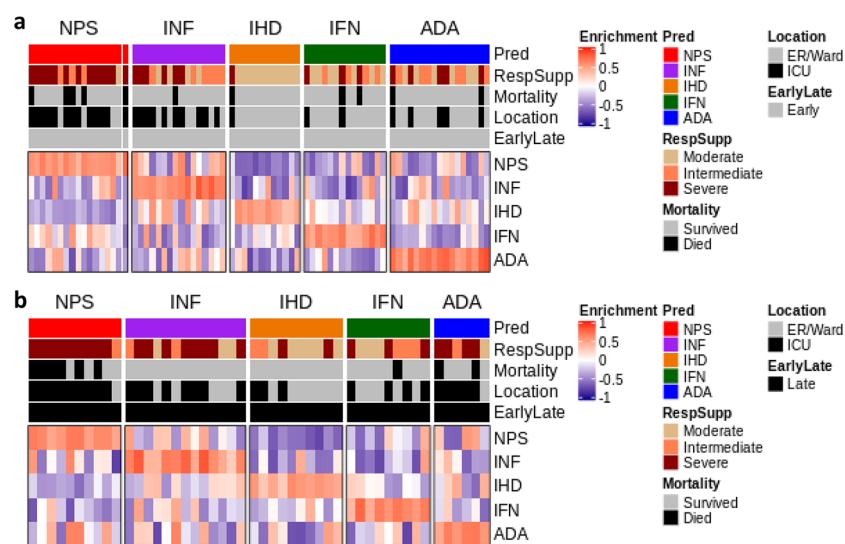


Figure 3. Heatmap depicting GSVA endotype signature enrichment scores and endotype classification COVID-19 patients. (a) Endotype classification of early sampled patients. (b) Endotype classification of late sampled patients. Endotype enrichment scores are indicated in the bottom five rows in each subfigure with red blocks indicating positive enrichment. Various metrics are indicated in the top five bars in each subgraph including endotype assignment, patient severity (as assessed by respiratory support requirement at sampling), impending mortality, and ICU admission, as well as Early (top subfigure) or Late (bottom subfigure) sampling. In both the early and late groups, the NPS endotype displayed the worst patient severity and highest mortality rates.

Glasgow Coma Score (p value = 0.017) (Table 3; Supplemental Table 2). The predicted NPS patients were primarily comprised of patients belonging to the severe respiratory support group (77%; 13/17) and also had the highest mortality, with 29.4% (5/17) of patients succumbing. The INF endotype also displayed a high number of patients belonging to the severe respiratory support group (37.5%; 6/16) in addition to the highest qSOFA scores (recorded early in hospitalization). In early sampled patients, the clinical trends of the predicted NPS and INF endotypes were generally consistent with the all-cause sepsis discovery cohort¹². However, a subgroup of more severely afflicted patients (higher severity and mortality) was also evident in the INF and ADA endotypes, which differed from trends observed in the smaller all-cause sepsis discovery and validation cohorts examined previously¹². For example, the patients predicted as part of the generally less severe ADA endotype, displayed high proportions of severe respiratory support patients (29.4%; 5/17).

In late sampled patients, the predicted NPS endotype patients again displayed the highest rates of mortality (60%) and ICU care (90%) (Supplemental Table 3). Beyond this observation, the patients assigned to other predicted endotypes showed less consistent clinical parameters relative to endotype assignment when compared to the early sampled patients (Table 3) and all-cause sepsis patients¹², possibly reflecting the fact that these endotypes were originally predicted at first clinical presentation (ER prior to ICU admission and any diagnosis). This was most apparent in ADA patients, who showed very high proportions requiring ICU care and severe respiratory support. Across early and late sampled patients, we reasoned that these inconsistencies could be attributed to mechanistic/gene-expression influences that were additional to the endotype signatures in COVID-19 patients (for example, Covid-specific gene expression, COVID-19 specific endotype signatures and/or mixed endotypes).

Comparing clinical data between endotypes indicated that sepsis endotypes might be useful for clinical risk stratification in COVID-19 disease. We next examined global gene expression differences between predicted endotypes to characterize highly dysregulated endotype-specific genes. It is reasonable to hypothesize that these genes might have regulatory functions and could be targeted for endotype-specific therapies. We specifically focused on characterizing the NPS and INF-predicted endotypes of the early sampled group, since detection and treatment would be most impactful in these endotypes and disease stage. DE gene expression analysis was performed comparing each endotype to all others (i.e., in a one vs. rest scheme) to identify global gene expression differences between endotypes. The NPS endotype displayed 2728 (1352 upregulated, 1376 down-regulated) unique DE genes while the INF endotype had 573 (520 upregulated, 53 down-regulated) unique DE genes.

These differences were explored in the context of protein:protein interaction (PPI) networks, specifically drawing from the International Molecular Exchange Consortium (ImeX) data, which includes the InnateDB database³³ (version 5.4) of manually curated immune interactions (Fig. 4). PPI networks reveal known direct-molecular/binding, metabolic or regulatory interactions between individual proteins, and thus represent function-based interactions in cells.

It was apparent that DE genes from the NPS endotype included both up- (red) and down- (green) regulated genes, consistent with the notion that the endotype captures concurrent neutrophil stimulation and immunosuppressive mechanisms (e.g., down-regulated IL-8, IL7R). Of most relevance was the identification of network hubs that indicate highly connected proteins/genes (shown by the size of the nodes/circles) and reflect their importance in the network as key signaling molecules that receive and integrate multiple biological signals. Importantly, these hubs reflect potential endotype-specific drug targets, which, when targeted, could disrupt the network structure and thus the endotype's pathophysiological basis of severity. Hubs of the NPS endotype network included peroxisome proliferator-activated receptor gamma (PPARG), synuclein alpha (SNCA), suppressor of cytokine signaling 3 (SOCS3), and C-X-C motif chemokine ligand 8/interleukin-8 (CXCL8/IL-8), among others. The INF endotype DE genes were primarily comprised of up-regulated genes. Hubs of the INF network included forkhead box O3 (FOXO3), galectin 3 (LGALS3), transglutaminase 2 (TGM2), ELAV Like RNA Binding Protein 1 (ELAV-1), and interferon-stimulated gene 15 (ISG15). Many of these can be targeted by existing drugs as compiled in Gene-Cards (<https://www.genecards.org/>) and thus are addressable with repurposed drugs (Supplemental Table 4).

Parameter	NPS (17/76)	INF (16/76)	IHD (12/76)	IFN (14/76)	ADA (17/76)	p value
Respiratory support group—moderate	5.9% (1/17)	25% (4/16)	91.7% (11/12)	50.0% (7/14)	35.3% (6/17)	0.000075
Receiving ICU care	76.5% (13/17)	62.5% (10/16)	8.3% (1/12)	14.3% (2/14)	29.4% (5/17)	0.00018
Respiratory support group—severe	76.5% (13/17)	37.5% (6/16)	8.3% (1/12)	21.4% (3/14)	29.4% (5/17)	0.0015
Glasgow coma score	13.3 ± 1.71 (7)	8 ± 2.86 (5)	15 ± 0 (5)	15 ± 0 (5)	15 ± 0 (9)	0.0174
P/F ratio	213.6 ± 19.82 (7)	175.2 ± 29.91 (5)	382.6 ± 21.36 (5)	274.4 ± 58.3 (5)	254.8 ± 31.26 (9)	0.022
qSOFA	0.6 ± 0.3 (7)	1.4 ± 0.4 (5)	0 ± 0 (5)	0.2 ± 0.2 (5)	0.4 ± 0.18 (8)	0.0454

Table 3. Clinical differences between predicted endotypes in the early sampled group. The mean value ± standard error is presented for numerical variables with the total available observations/patient numbers recorded in brackets. Categorical variables are presented as % positive (relative to total available observations). Significance was assessed using a Kruskal–Wallis and Chi Square Test for numeric and categorical parameters, respectively. The P/F ratio is the arterial pO₂ (“P”) divided by the FIO₂ (“F”) with a P/F Ratio less than 300 indicating acute respiratory failure³². A full table of clinical parameters assessed, including non-significant parameters, is provided in Supplemental Table 2.

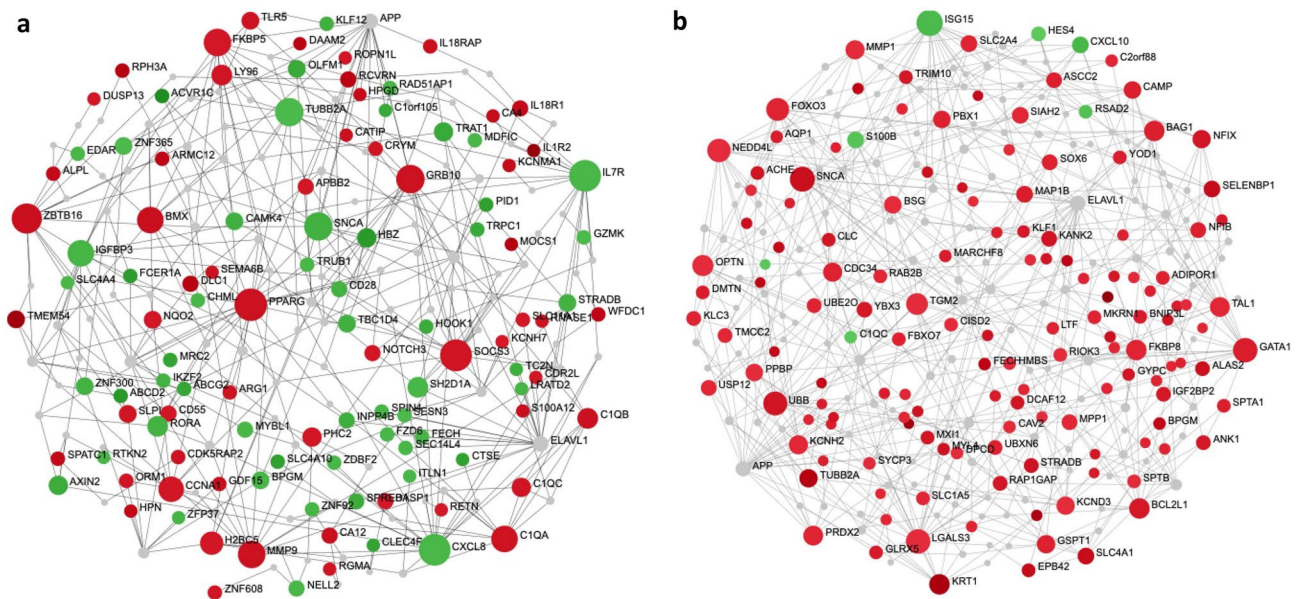


Figure 4. Function-based, minimally connected first order protein:protein interaction networks of the top endotype specific DE genes. **(a)** NPS endotype network. **(b)** INF endotype network. The top 200 endotype-specific DE genes (by absolute value fold changes) were input into NetworkAnalyst⁵⁴. Each DE gene set formed a well-connected minimum connected first-order network, indicating that the genes involved are functionally related and likely collectively regulate or play key roles in one or more related biological mechanisms. Red and green nodes are genes with increased and decreased expression specific to the endotype, respectively, while grey nodes are interconnecting first-order interaction nodes. The size of the nodes indicates their connectivity (hub degree) within the network, i.e., how well interconnected any given node is to other nodes in the Network (NB. hubs are key molecules in signalling since they are highly interconnected; they receive and integrate multiple signals and pass them on to downstream nodes). Lines represent edges that indicate known (experimentally-determined) protein:protein interactions derived from www.innatedb.ca (version 5.4).

Discussion

In the current COVID-19 pandemic, severely affected patients have been proposed to suffer from sepsis^{34,35}. However, given the enormous heterogeneity of sepsis, it has not been clear if this is a distinct type of sepsis or mechanistically overlapping with all-cause sepsis. Therefore, we analyzed the gene expression profiles of COVID-19 patients to examine dominant gene expression responses. We show that all-cause sepsis signatures of severity, mortality, and mechanisms/endotypes are dysregulated in severe COVID-19 patients, indicating, based on molecular responses, that severe COVID-19 is a form of all-cause sepsis. This is significant because it implies severe COVID-19 can be predicted using the same biomarkers and gene expression signatures useful in all-cause sepsis patients. Furthermore, by identifying their endotypes, Covid-19 patients can be potentially treated with immunomodulatory therapies that correct dysfunctional immune processes specific to a particular endotype, as predicted in Supplemental Table 4. Given several reports indicating that antiviral therapies display no marked efficacy in reducing mortality in COVID-19 patients³⁶, this is especially relevant.

We initially identified COVID-19 specific mechanisms that were dysregulated between severity groups (Fig. 1). These results showed that in addition to sepsis-related gene expression (e.g., Neutrophil degranulation, various inflammatory hallmarks), COVID-19 positive patients displayed significant dysregulation of interferon and heme metabolism response that are not as obvious in all-cause sepsis patients. The up-regulation of Type I and II interferon-related pathways in severe COVID-19 has been observed in many previous studies exploring COVID-19 infections^{37,38}, and is generally observed in viral infections. COVID-19 patients have also been shown to have excessive and dysfunctional neutrophils^{39,40}, although it is worth mentioning that this is a general feature of patients with more severe cases of sepsis⁴¹.

Heme metabolism was also strongly dysregulated when comparing patients who died vs. survived. In all-cause sepsis, it has been shown that sepsis survivors display a signature of heme biosynthesis, wherein up-regulated heme synthesis prevents further tissue damage from infection⁴². Other studies have shown that severe sepsis patients fail to remove excessive heme leading to tissue damage^{28,43}. Some literature exploring heme metabolism in COVID-19 patients indicates that SARS-CoV-2 proteins directly bind hemoglobin inhibiting their metabolism, although this is debated since it has only been posited through in silico experiments^{44,45}.

The CR and Organ Dysfunction signatures, which have been shown to predict impending severe sepsis^{11,12}, performed well in discriminating between severe, intermediate, and moderate patients in early and late sampled patients with COVID-19. Interestingly, the CR signature, which was originally derived from treating PBMCs twice with lipopolysaccharide (a bacterial agonist) yielding an immune tolerance phenotype, was strongly able to discriminate between severity groups. It is unknown whether signatures of this form of immune tolerance are the same irrespective of bacterial or viral stimuli although both bacterial and viral-like molecules can induce such

tolerance^{46,47}. Nonetheless, our results clearly indicate that the CR signature is strongly associated with severity due to viral infections and thus reflects an important mechanism (monocyte reprogramming to a non-responsive phenotype) of COVID-19 sepsis. Interestingly, a recent publication showed that the stimulation of human blood cells with SARS-CoV-2 proteins induced hallmarks of CR, including reduced proinflammatory cytokine production and T cell proliferation, and enhancement of phagocytosis and tissue remodeling⁴⁸. Furthermore, there have been other studies which have shown through bioinformatic analysis and mathematical models that related biological responses dysregulated during severe Covid-19 infection overlap with that of all-cause sepsis^{38,49,50}. The Mortality signature also captured 30-day mortality quite well, indicating that similar mechanisms arise during severe COVID-19 infection and all-cause sepsis when death is impending.

We also explored endotypes within the COVID-19 gene expression data. Endotypes reflect a more recent paradigm to characterize the sepsis syndrome, whereby each endotype captures distinct mechanisms that an individual patient could express. Accordingly, we predicted endotype status among early and late sampled COVID-19 patients. Predictions generated by GSVA enrichment scores indicated that all-cause sepsis endotypes can, to a great extent, explain the heterogeneity of severe COVID-19 disease. Additional complexity was also observed (Fig. 3), perhaps due to COVID-19-driven gene expression and the somewhat varied time of sampling relative to hospital admission. Thus sepsis and/or COVID-19 signatures may exhibit somewhat different behaviour in the earlier and later stages of the disease, which is consistent with previous studies by our group indicating that gene expression trajectories change over two time points taken seven days apart in severe COVID-19 patients (An et al., manuscript submitted).

Results were most consistent with the expected behaviour of endotypes in the early sampled patients, which is reasonable considering that the endotypes were identified in ER patients at first clinical presentation (within 2 h of admission). Specifically, the predicted NPS endotype patients exhibited the highest mortality and disease severity based on respiratory support requirements. Sweeney et al.⁵¹ presented an analogous result when applying their three-endotype classifier to COVID-19 patients. That study showed that, based on the expression of 33 biomarker genes (5–17 per endotype), the genes were also expressed in COVID-19 patients, and the defining clinical trends for each endotype were mostly maintained. It is worth noting however that our 5 endotype classifier not only classifies patients into mechanistically and to some extent clinically distinct groups, but each of these endotypes can be accurately classified by expression of a diverse set of gene pairs¹².

In conclusion, we present compelling evidence that COVID-19 is a type of sepsis based on the dysregulation of specific gene expression signatures, with unique COVID-19 specific immune responses also contributing to patient severity. When exploring the expression of previously-described¹² all-cause sepsis signatures in a broad spectrum of COVID-19 patients, severely ill patients displayed significant dysregulation of these genes. Furthermore, several COVID-19 gene expression responses have known roles in sepsis pathogenesis, including various inflammatory pathways, neutrophil degranulation, interferon α/β signaling and heme metabolism. This can provide prognostic criteria as well as a precision medicine approach to characterize (and possibly treat) the disease. In conclusion, future opportunities for the clinical management of Covid-19 patients can draw from existing and under-development prognostic tools relevant to all-cause sepsis signatures and therapeutic options that emphasize address a dysregulated host immune response.

Data availability

Sequencing data and patient details are available at NCBI GEO (Accession Number GSE221234). Other materials, including study protocol, statistical analysis plan, and code is available to anyone who requests. These requests should be directed to REW Hancock (bob@hancocklab.com).

Received: 9 August 2022; Accepted: 16 January 2023

Published online: 23 January 2023

References

- Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet* **395**(10229), 1054–1062 (2020).
- Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* **395**(10223), 507–513 (2020).
- Bhatraju, P. K. *et al.* COVID-19 in critically ill patients in the Seattle region—case series. *N. Engl. J. Med.* **382**(21), 2012–2022 (2020).
- Weerahandi, H. *et al.* Post-discharge health status and symptoms in patients with severe COVID-19. *J. Gen. Int. Med.* **36**(3), 738–745 (2021).
- Li, H. *et al.* SARS-CoV-2 and viral sepsis: Observations and hypotheses. *Lancet* **395**(10235), 1517–1520 (2020).
- Sharifpour, E. *et al.* Evaluation of bacterial co-infections of the respiratory tract in COVID-19 patients admitted to ICU. *BMC Infect. Dis.* **20**(1), 1–7 (2020).
- Shafraan, N. *et al.* Secondary bacterial infection in COVID-19 patients is a stronger predictor for death compared to influenza patients. *Sci. Rep.* **11**(1), 1–8 (2021).
- Vincent, J. L. COVID-19: It's all about sepsis. *Future Microbiol.* **16**(3), 131–133 (2021).
- Leligdowicz, A. & Matthay, M. A. Heterogeneity in sepsis: New biological evidence with clinical applications. *Crit. Care* **23**(1), 1–8 (2019).
- Stanski, N. L. & Wong, H. R. Prognostic and predictive enrichment in sepsis. *Nat. Rev. Nephrol.* **16**(1), 20–31 (2020).
- Pena, O. M. *et al.* An endotoxin tolerance signature predicts sepsis and organ dysfunction at initial clinical presentation. *EBioMed* **1**(1), 64–71 (2014).
- Baghela, A. *et al.* Predicting sepsis severity at first clinical presentation: The role of endotypes and mechanistic signatures. *EBioMed* **75**, 103776 (2022).
- Schultze, J. L. & Aschenbrenner, A. C. COVID-19 and the human innate immune system. *Cell* **184**(7), 1671–1692 (2021).
- Tremblay, K. *et al.* The Biobanque québécoise de la COVID-19 (BQC19)—A cohort to prospectively study the clinical and biological determinants of COVID-19 clinical trajectories. *PLoS One* **16**(5), e0245031 (2021).

15. Skogholt, A. H. *et al.* Gene expression differences between PAXgene and Tempus blood RNA tubes are highly reproducible between independent samples and biobanks. *BMC Res. Notes* **10**(1), 1–2 (2017).
16. Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
17. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**(19), 3047–3048 (2016).
18. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**(1), 15–21 (2013).
19. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**(2), 166–169 (2015).
20. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**(12), 1–21 (2014).
21. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**(6), 882–883 (2012).
22. Fabregat, A. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **46**(D1), D649–D655 (2018).
23. Liberzon, A. *et al.* The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**(6), 417–425 (2015).
24. Pena, O. M., Pistolic, J., Raj, D., Fjell, C. D. & Hancock, R. E. Endotoxin tolerance represents a distinctive state of alternative polarization (M2) in human mononuclear cells. *J. Immunol.* **186**(12), 7243–7254 (2011).
25. Wu, D. *et al.* ROAST: Rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**(17), 2176–2182 (2010).
26. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7), e47 (2015).
27. Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**(1), 1–5 (2013).
28. Olonisakin, T. F. *et al.* Stressed erythrophagocytosis induces immunosuppression during sepsis through heme-mediated STAT1 dysregulation. *J. Clin. Investig.* **131**(1), 25 (2021).
29. Kangelaris, K. N. *et al.* Increased expression of neutrophil-related genes in patients with early sepsis-induced ARDS. *Am. J. Physiol. Lung Cell Mol. Physiol.* **308**(11), L1102–L1113 (2015).
30. Liu, X. *et al.* Prognostic significance of neutrophil-to-lymphocyte ratio in patients with sepsis: A prospective observational study. *Mediat. Inflamm.* **20**, 16 (2016).
31. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. *InCancer Syst. Biol.* **20**, 243–259 (2018).
32. Esper, A. M. *et al.* The role of infection and comorbidity: Factors that influence disparities in sepsis. *Crit. Care Med.* **34**(10), 2576 (2006).
33. Breuer, K. *et al.* InnateDB: Systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* **41**(D1), D1228–D1233 (2006).
34. Beltrán-García, J. *et al.* Sepsis and coronavirus disease 2019: Common features and anti-inflammatory therapeutic approaches. *Crit. Care Med.* **48**(12), 1841–1844 (2020).
35. Prescott, H. C. & Girard, T. D. Recovery from severe COVID-19: Leveraging the lessons of survival from sepsis. *JAMA* **324**(8), 739–740 (2020).
36. Veghivinti, C. T. *et al.* Efficacy of antiviral therapies for COVID-19: A systematic review of randomized controlled trials. *BMC Infect. Dis.* **22**(1), 1–45 (2022).
37. Mesev, E. V., LeDesma, R. A. & Ploss, A. Decoding type I and III interferon signalling during viral infection. *Nat. Microbiol.* **4**, 914–924 (2019).
38. Wilk, A. J. *et al.* A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat. Med.* **26**(7), 1070–1076 (2020).
39. Aschenbrenner, A. C. *et al.* Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. *Genome Med.* **13**(1), 1–25 (2021).
40. Arcanjo, A. *et al.* The emerging role of neutrophil extracellular traps in severe acute respiratory syndrome coronavirus 2 (COVID-19). *Sci. Rep.* **10**(1), 1 (2020).
41. Sônego, F. *et al.* Paradoxical roles of the neutrophil in sepsis: Protective and deleterious. *Front. Immunol.* **7**, 155 (2016).
42. Leite, G. G., Scicluna, B. P., van Der Poll, T. & Salomão, R. Genetic signature related to heme-hemoglobin metabolism pathway in sepsis secondary to pneumonia. *NPJ Syst. Biol.* **5**(1), 1–9 (2019).
43. Larsen, R. *et al.* A central role for free heme in the pathogenesis of severe sepsis. *Sci. Trans. Med.* **2**(51), 5171 (2010).
44. Cavezzi, A., Troiani, E. & Corrao, S. COVID-19: Hemoglobin, iron, and hypoxia beyond inflammation. A narrative review. *Clin. Pract.* **10**(2), 24–30 (2020).
45. DeMartino, A. W. *et al.* No evidence of hemoglobin damage by SARS-CoV-2 infection. *Haematologica* **105**(12), 2769 (2020).
46. Sato, S. *et al.* A variety of microbial components induce tolerance to lipopolysaccharide by differentially affecting MyD88-dependent and-independent pathways. *Int. Immunol.* **14**(7), 783–791 (2002).
47. Butcher, S. K., O’Carroll, C. E., Wells, C. A. & Carmody, R. J. Toll-like receptors drive specific patterns of tolerance and training on restimulation of macrophages. *Front. Immunol.* **9**, 933 (2018).
48. Avendaño-Ortiz, J. *et al.* SARS-CoV-2 proteins induce endotoxin tolerance hallmarks: A demonstration in patients with COVID-19. *J. Immunol.* **207**(1), 162–174 (2021).
49. Day, J. D. *et al.* Divergent COVID-19 disease trajectories predicted by a DAMP-centered immune network model. *Front. Immunol.* **12**, 4383 (2021).
50. Laing, A. G. *et al.* A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat. Med.* **26**(10), 1623–1635 (2020).
51. Sweeney, T. E. *et al.* Validation of inflammopathic, adaptive, and coagulopathic sepsis endotypes in coronavirus disease 2019. *Crit. Care Med.* **49**(2), e170–e178 (2021).
52. Brown, S. M. *et al.* Non-linear imputation of PaO₂/FIO₂ from SpO₂/FIO₂ among mechanically ventilated patients in the intensive care unit: A prospective, observational study. *Crit. Care Med.* **45**(8), 1317 (2017).
53. Zhao, S., Li, C. I., Guo, Y., Sheng, Q. & Shyr, Y. RnaSeqSampleSize: Real data based sample size estimation for RNA sequencing. *BMC Bioinform.* **19**(1), 1–8 (2018).
54. Zhou, G. *et al.* NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **47**(W1), W234–W241 (2019).

Acknowledgements

Funding from the Canadian Institutes for Health Research (CIHR) COVID-19 Rapid Research. Funding to RH, AL and RL, CIHR FDN-154287 to RH, and CIHR VR2-173203 to DEK is gratefully acknowledged. RH held a UBC Killam Professorship. AA was funded by a Canada Graduate Scholarships Doctoral (CGS-D) program. This

work was made possible through open sharing of data and sample from the Biobanque Québécoise COVID-19, funded by the Fonds de recherche du Québec—Santé, Génome Québec and the Public Health Agency of Canada.

Author contributions

R.H. and R.L. conceived the study. A.B., A.A., and R.H. contributed to the study design. A.B. performed bioinformatics analysis. A.B. and R.H. drafted the manuscript. A.B., A.A., P.Z., E.A., J.G., T.B., G.F., A.L., and R.H. contributed to interpretation of data. A.B., G.C.F., and R.H. contributed to statistical analysis. E.R., D.K., and R.L. coordinated and were directly involved in, sample and patient metadata collection in hospitals. All authors edited the manuscript. A.B., A.A., E.A., and A.L. verified the quality and accuracy of sequencing and clinical data. R.H., A.L. and R.L. were responsible for obtaining funding and R.H. led the study. All authors have read and approved the final version of the manuscript.

Competing interests

RH has filed the CR signature for patent protection and licenced this to Asep Medical, a Vancouver company in which he has a significant ownership position. RH, GCF, and AB have filed the all-cause sepsis endotype signatures for patent protection. 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/s41598-023-28259-y>.

Correspondence and requests for materials should be addressed to R.E.W.H.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023