# Microbial signatures in the lower airways of mechanically ventilated COVID19 patients associated with poor clinical outcome

*A full list of authors and affiliations appears at the end of the article.*

## Abstract

Respiratory failure is associated with increased mortality in COVID-19 patients. There are no validated lower airway biomarkers to predict clinical outcome. We investigated whether bacterial respiratory infections were associated with poor clinical outcome of COVID-19 in a prospective, observational cohort of 589 critically ill adults, all of whom required mechanical ventilation. For a subset of 142 patients who underwent bronchoscopy we quantified SARS-CoV-2 viral load, analysed the lower respiratory tract microbiome using metagenomics and metatranscriptomics and profiled the host immune response . Acquisition of a hospital-acquired respiratory pathogen was not associated with fatal outcome. Poor clinical outcome was associated with lower airway enrichment with an oral commensal (*Mycoplasma salivarium*). Increased SARS-CoV-2 abundance, low anti-SARS-CoV-2 antibody response and a distinct host transcriptome profile of the lower airways were most predictive of mortality. Our data provide evidence that secondary respiratory infections do not drive mortality in COVID-19 and clinical management strategies should prioritize reducing viral replication and maximizing host responses to SARS-CoV-2.

## Introduction

The earliest known case of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection causing coronavirus virus disease (COVID-19) is thought to have occurred on November 17, 2019[1]. As of August 3, 2021, 198.7 million confirmed cases of COVID-19 and 4.2 million deaths have been reported worldwide[2]. As the global scientific community rallied in a concerted effort to understand SARS-CoV-2 infections, our background knowledge is rooted in previous experience with the related zoonotic betacoronaviruses Middle East Respiratory Syndrome coronavirus (MERS-CoV) and SARS-CoV-1 that have caused severe pneumonia with 34.4% and 9% case fatality, respectively[3]. As observed for these related coronaviruses, SARS-CoV-2 infection can result in an uncontrolled

*Corresponding Authors: Elodie Ghedin (elodie.ghedin@nih.gov) and Leopoldo N. Segal, (Leopoldo.Segal@nyumc.org).

inflammatory response[4] leading to acute respiratory distress syndrome (ARDS) and multi-organ failure, both associated with increased mortality. While a large proportion of the SARS-CoV-2 infected population is asymptomatic or experiences mild illness, a substantial number of individuals will develop severe disease and require hospitalization, with some progressing to respiratory failure and death.

Mortality in other viral pandemics, such as the 1918 H1N1 and 2009 H1N1 influenza pandemics, has been attributed in part to bacterial co-infection or super-infection[5,6]. To determine if this is also the case for COVID-19, we can use next generation sequencing (NGS) to probe the complexity of the microbial environment (including RNA and DNA viruses, bacteria and fungi) and how the host (human) responds to infection. Recent studies have used this approach to uncover microbial signatures in patients with ARDS.[7,8] Increased bacterial burden and the presence of gut-associated bacteria in the lung were shown to worsen outcomes in these critically ill patients[7,9], highlighting the potential role of the lung microbiome in predicting outcomes in ARDS. In a recent study using whole genome sequencing to profile the gut microbiome of 69 patients from Hong Kong, investigators identified an increased abundance of opportunistic fungal pathogens among patients with confirmed COVID-19[10]. While there is emerging interest in understanding the microbial environment in patients with SARS-CoV-2 infections, few studies have attempted to characterize this at the primary site of disease activity: the lower airways[11,12].

In this investigation, we characterized the lung microbiome and lower airway markers of host immunity in a cohort of hospitalized COVID-19 patients. While we did not find that isolation of a secondary respiratory pathogen was associated with prolonged mechanical ventilation (>28 days) or fatal outcome, we did identify critical microbial signatures—characterized by enrichment of oral commensals, high SARS-CoV-2 load, and decreased anti-SARS-CoV-2 IgG response—associated with fatal outcome, suggesting a need for more targeted antiviral therapeutic approaches for the care of critically ill COVID19 patients.

## Results

### Cohort description

From March 3rd to June 18th 2020, a total of 589 patients with laboratory-confirmed SARS-CoV-2 infection were admitted to the intensive care units of two academic medical centers of NYU Langone Health in New York (Long Island and Manhattan) and required invasive mechanical ventilation (MV) (see Supplementary Text and Supplementary Tables 1 and 2). This included a subset of 142 patients from the Manhattan campus who underwent bronchoscopy for airway clearance and/or tracheostomy from which we collected and processed lower airway (BAL) samples for this investigation (Extended Data Fig. 1). Table 1 shows demographics and clinical characteristics of the 142 patients who underwent bronchoscopy divided into three clinical outcomes: survivors with 28 Days on MV; survivors with >28 Days on MV; and deceased. The median post admission follow-up time was 232 days (CI=226–237 days). Patients within the bronchoscopy cohort had a higher overall survival than the rest of the NYU COVID-19 cohort since most critically ill patients were not eligible for bronchoscopy or tracheostomy.

Among the factors associated with clinical outcome within the bronchoscopy cohort, patients who survived were more commonly placed on veno-venous extracorporeal membrane oxygenation (ECMO) whereas patients who died had frequently required dialysis (Table 1). These trends were also observed across the whole NYU cohort. Neither hydroxychloroquine nor azithromycin were significantly associated with clinical outcome. However, patients who survived were more frequently treated with the combination antibiotic piperacillin/tazobactam.

Within the first 48hrs from admission, respiratory bacterial cultures were rarely obtained (n=70/589, 12%) with very few positive results (n=12, 17%). Blood cultures were more commonly obtained (n=353/589, 60%) but the rate of bacterial culture positivity was much lower (n=5, 1.4%). These data support that community acquired bacterial co-infection was not a common presentation among critically ill COVID-19 patients.

We evaluated whether respiratory or blood culture results obtained as per clinical standard of care were associated with clinical outcome. Risk analyses for the culture results during hospitalization for the whole cohort (n=589) demonstrated that bacterial culture positivity with suspected pathogens—excluding possible contaminants such as oral microbes in respiratory samples—was not associated with increased odds of dying but was associated with prolonged mechanical ventilation in the surviving patients (Figure 1). Since length of stay could potentially affect these results (patients who died could have a shorter hospitalization, and therefore may have had fewer specimens collected for cultures), we repeated the analysis using culture data obtained during the first two weeks of hospitalization. This analysis showed that bacterial pathogen culture positivity (both respiratory and blood) during the early period of hospitalization was not associated with worse outcome (Figure 1 and Supplementary Table 3). Interestingly, identification of oral bacteria in respiratory culture, commonly regarded as procedural contaminants, was associated with higher odds of prolonged mechanical ventilation (>28 days) among survivors. Similar trends were noted when analysis was performed on subjects from NYU LI and NYU Manhattan separately, or for the bronchoscopy cohort (Supplementary Table 2). Among the bronchoscopy cohort, there was no statistically significant association between culture results and clinical outcome, but there was a trend towards an increased rate of positive respiratory cultures for *Staphylococcus aureus* (including MRSA), *S. epidermidis*, and *Klebsiella pneumoniae* in the survival groups (Table 1). These data suggest that in critically ill patients with COVID-19 requiring MV, in whom broad spectrum antimicrobials were frequently used, hospital isolation of a secondary respiratory bacterial pathogen is not associated with worse clinical outcome.

### SARS-CoV-2 load in the lower airways

Using bronchoscopy samples from 142 patients we evaluated the viral load by rRT-PCR for the SARS-CoV-2 N gene, adjusted by levels of human ribosomal protein (RP). Of note, the majority of samples were largely obtained in the second week of hospitalization (Table 1, median[IQR] = 10[6–14], 13[8–16], and 13[8–16] for the 28-days MV, >28-days MV, and deceased groups, respectively, p=ns). Paired analysis of upper airways (UA) and BAL samples revealed that, while there was a positive association between SARS-CoV-2 viral

load of the paired samples (rho = 0.60, p<0.0001), there was a subset of subjects (21%) for which the viral load was greater in the BAL than in the supraglottic area, indicating topographical differences in SARS-CoV-2 replication (Figure 2a). Importantly, while the SARS-CoV-2 viral load in the UA samples was not associated with clinical outcome (Supplementary Fig. 1), patients who died had higher viral load in their lower airways than patients who survived (Figure 2b). Several studies have explored the relationship between SARS-CoV-2 viral load and mortality[13–18]. In a large cohort of 1145 patients with confirmed SARS-CoV-2, viral load measured in nasopharyngeal swab samples was found to be significantly associated with mortality, even after adjusting for age, sex, race and several co-morbidities[18]. Similar results were found in a cohort of patients in New York City with or without cancer, where in-hospital mortality was significantly associated with a high SARS-CoV-2 viral load in the UA[17].

We then evaluated virus replication in BAL samples by measuring levels of subgenomic RNA (sgRNA) targeting the E gene of SARS-CoV-2. This mRNA is only transcribed inside infected cells and is not packaged into virus particles; thus, its presence is indicative of recent virus replication[19–21]. In BAL, levels of sgRNA correlated with viral load as estimated by rRT-PCR for the SARS-CoV-2 N gene (Figure 2c) and the highest percentage of measurable sgRNA was in the deceased group followed by the   28-days MV group, and the >28-days MV group (17,7%, 11.5%, and 3.7%, respectively, chi-square p=0.028 for the comparison deceased vs. >28-days MV group). Thus, while in most cases levels of sgRNA were not measurable in BAL suggesting that no active virus replication was ongoing in the lower airways of COVID-19 patients at the time of bronchoscopy (overall median[IQR] = 12[7–16] days from hospitalization), the lower airway viral burden, as estimated by rRT-PCR, is associated with mortality in critically ill COVID-19 patients.

### Microbial community structure of the lower & upper airways

Considering the bacterial species and the viral loads identified in the BAL and UA of this cohort and their association with outcomes, we profiled in detail their viral and microbial composition. Microbial communities were evaluated using parallel datasets of RNA and DNA sequencing from 118 COVID-19 patients with BAL samples that passed appropriate quality control and a subset of paired 64 UA samples, along with background bronchoscope controls.

Given the low biomass of BAL samples in the metatranscriptome data, we first identified taxa as probable contaminants by comparing the relative abundance between background bronchoscope and BAL samples (Extended Data Fig. 2a and Supplementary Table 4). However, we did not remove any taxa identified as probable contaminants from subsequent analyses. A comparison of the microbial community complexity captured in these data, determined using the Shannon diversity Index, showed there was significantly lower α diversity in the BAL samples than in the UA and background controls (Extended Data Fig. 3a). Similarly, β diversity analysis based on the Bray Curtis Dissimilarity index indicated that the microbial composition of the lower airways was distinct from the UA and background controls (Extended Data Fig. 3b, PERMANOVA p<0.01). Sequence reads indicated a much higher relative abundance of SARS-CoV-2 in the lower airways than in

the UA for this cohort (Extended Data Fig. 3c). Comparisons of the most dominant bacterial and fungal taxa that were functionally active showed that *S. epidermidis*, *Mycoplasma salivarium*, *S. aureus*, *Prevotella oris*, and *Candida albicans*, many often considered oral commensals, were present in both UA and BAL samples (Extended Data Fig. 3c). Interestingly, the lytic phage Proteus virus Isfahan, known to be active against biofilms of *Proteus mirabilis*[22], was found to be highly transcriptionally active in the BAL.

As with the metatranscriptome data, we first identified taxa as probable contaminants in the metagenome data but these were not removed for subsequent analyses (Extended Data Fig. 2b). Both α and β diversity analyses of the metagenome support distinct microbial community features in the lower airways as compared with the UA and background controls (Extended Data Fig. 4a, 4b). Interestingly, *S. epidermis* ranked as the most highly functional taxon in both BAL and UA, based on RNAseq reads (Extended Data Fig. 3c), was 33rd in relative abundance in the BAL DNAseq data but present at very high relative abundance in the UA (ranked #3). These data suggest that microbes that colonize the UA and the skin were common in the lower airways in this cohort of COVID-19 patients requiring invasive MV.

**Airway microbiota are associated with clinical outcomes—**Consistent with the SARS-CoV-2 viral load assessed by RT-PCR, differential expression analysis (DESeq) of the RNA virome identified SARS-CoV-2 as being enriched in the deceased group, as compared with both ≤28-days and >28-days MV groups (fold change >5, Figure 2d). Cox proportional hazards modeling supports that enrichment with SARS-CoV-2 was associated with increased risk for death (HR 1.33, 95% CI= 1.07–1.67, pvalue=0.011, FDR adjusted pvalue=0.06; Supplementary Table 5).

The phage metatranscriptome α and β diversity was similar across the clinical outcome groups. There were, however, various taxonomic differences at the RNA level with enrichment of *Staphylococcus* phages CNPx in the deceased and >28-day MV groups when compared with the ≤28-day MV group (Figure 2e). Differential expression from two other *Staphylococcus* phages was also observed in the >28-days MV group as compared with the ≤28-days MV group (Figure 2e). None of the described taxa were identified as possible contaminants (Supplementary Table 4).

## Oral commensals and poor clinical outcome

We evaluated the overall bacterial load by quantitative PCR, targeting the 16S rRNA gene. As expected, the bacterial load in the BAL was several folds lower than in the UA but clearly higher than the background bronchoscope control (Supplementary Fig. 2). Patients who died had higher total bacterial load in their BAL than patients who survived (Figure 3a).

While no statistically significant differences were noted in α or β diversity across clinical outcome groups (Figure 3b–c), several differences were noted when differential enrichment was evaluated using DESeq. For the comparisons made across the clinical outcome groups we focused on consistent signatures identified in the lower airway metagenome and metatranscriptome. Coherence of differentially enriched taxa was determined by gene set enrichment analysis (GSEA) (Figure 3d) and directionality of enrichment between the two

datasets was evaluated (Figure 3e). Among the most abundant taxa, the oral commensal *M. salivarum* was enriched in the deceased and >28-days MV groups as compared with the ≤28-days MV group. In contrast, a different oral commensal, *Prevotella oris*, was enriched in the ≤28-days MV group as compared with the deceased and >28-days MV groups. From previous data published by us, enrichment of the lower airway microbiota with oral commensals was seen to be associated with a pro-inflammatory state in several diseases, including lung cancer[23,24] and non-tuberculosis mycobacterium related bronchiectasis[25]. The data in this analysis support that oral commensals are frequently found in the lower airways of critically ill COVID-19 patients and that differences between groups could be due to differential microbial pressures related to host factors. Interestingly, most of the statistically significant taxa were identified in the metatranscriptome rather than in the metagenome data, with only *P. oris* identified in both datasets. None of the described taxa were identified as possible contaminants (Supplementary Table 4).

For the fungal data, there were no statistically significant differences in α or β diversity identified between clinical outcome groups in the metagenome or the metatranscriptome data (Extended Data Fig. 5a and 5c). However, in the metagenome data, we identified *Candida glabrata* enriched in the deceased group as compared with the ≤28-days MV and the >28-days MV groups but this was not consistent in the metatranscriptome data (Extended Data Fig. 5b and 5d).

**Microbial functional profile and poor outcome**

We used the gene annotation of the DNAseq and RNAseq data to profile the microbial functional potential of the BAL samples. For the comparisons made across the clinical outcome groups, we focused on consistent functional signatures identified in the lower airway metagenome and metatranscriptome. Coherence of differentially enriched functions was determined using GSEA (Extended Data Fig 6a) and directionality of enrichment was also evaluated (Extended Data Fig 6b). Overall, there was coherence of directionality between the metranscriptomics and metagenomics datasets for the comparisons between deceased vs ≤28-days MV, and >28-days MV vs ≤28-days MV groups. Interestingly, statistically significant differences were only noted in the metatranscriptome data and not in the metagenome data, suggesting that functional activation of microbes can provide further insights into the lower airway microbial environment of patients with worst outcome. Among the top differentially expressed pathways in the poor outcome groups were glycosylases, oxidoreductase activity, transporters, and the two-component system, which is used by bacteria and fungi for signaling. A specific analysis of antibiotic resistance genes showed that there was significant gene enrichment and expression of biocide resistance in the deceased group as compared to the two other MV groups (Extended Data Fig. 7). There was also significant expression of genes resistant to trimethoprim and phenolic compound, as well as multi drug resistance in the deceased group as compared to the ≤28-days MV group. Presence of the resistance gene against trimethoprim was not significantly associated with prior exposure to the drug. However, only 7 patients received this drug before sample collection. These differences may indicate important functional differences leading to a different metabolic environment in the lower airways that could impact host

immune responses. It could also be representative of differences in microbial pressure in patients with higher viral loads and different inflammatory environments.

### Adaptive & innate immune responses to SARS-CoV-2

To evaluate the host immune response to SARS-CoV-2 infection, we first measured levels of anti-Spike and anti-RBD (receptor binding domain) antibodies in BAL samples. For both anti-Spike and anti-RBD immunoglobulins, levels of IgG, IgA and IgM were several logs higher than levels found in BAL samples from non-SARS-CoV-2 infected patients. Importantly, IgG levels of anti-Spike and anti-RBD were significantly lower in the deceased group as compared to the levels found in patients who survived (Figure 4a and Extended Data Fig. 8a–c, $p < 0.05$). Prior investigations have suggested that IgA levels are a key driver of neutralization in the mucosa[26–28]. The differences noted in the current investigation in the IgG pools are intriguing and future work investigating the antibodies generated during SARS-CoV-2 infections will be essential. Additionally, a neutralization assay performed using BAL fluid showed varying levels of neutralization across all samples (as estimated by EC50) but no statistically significant differences between the clinical outcome groups (Extended Data Fig. 8d).

Host transcriptome analyses of BAL samples showed significant differences across clinical outcome groups based on β diversity composition (Extended Data Fig. 9). We identified multiple differentially expressed genes across the clinical outcome groups (Extended Data Fig. 9b-d). First, we noted that the lower airway transcriptomes showed downregulation of heavy constant of IgG (IGHG3), and heavy constant of IgA (IGHA1) genes in those with worse clinical outcome (Supplementary Table 6). We then used IPA (Ingenuity Pathway Analysis) to summarize differentially expressed genes across the three clinical outcome groups (Figure 4b). The sirtuin Signaling Pathway (a pathway known to be involved in aging, gluconeogenesis/lipogenesis, and host defense against viruses)[29] and the ferroptosis pathway (an iron-dependent form of regulated cell death present in bronchial epithelium)[30,31] were both upregulated in those with worse outcome. Interestingly, there is evidence to support that STAT3[32] and ACSL4[33] alleviated ferroptosis-mediated acute lung injury dysregulation, which are both down-regulated in COVID-19 patients with worse clinical outcome. While this may reflect the host response to viral infection, other differences in the transcriptomic data showed downregulation of mitochondrial oxidative phosphorylation, HIF1α, STAT3, and Phospholipase C Signaling. Additional canonical signaling pathways, including insulin secretion, multiple Inositol related pathways, noradrenaline/adrenaline degradation signaling, and xenobiotic related metabolism were significantly downregulated when comparing the >28-days MV vs.  28-days MV groups. There is evidence that in the neonatal lung, inositol related components exert an anti-inflammatory effect and can prevent acute lung injury[34,35].

To determine if the abundance of immune cells varies between different clinical outcome groups, we estimated cell type abundance from the host transcriptome with computational cell type quantification methods, including a deconvolution approach implemented in CIBERSORTx[36] and a cell type signature enrichment approach implemented in xCell[37]. As reported recently in other studies[38], among the cell types detected in the BAL samples

we observed a consistent enrichment of mast cells and neutrophils in the >28-days MV and deceased groups compared with the 28-days MV group (Figure 4c and Supplementary Table 7). We also identified significantly higher inflammatory macrophages (M1), innate T-cells and memory T-cells (CCR7+) among subjects with worse clinical outcome.

### Cross-kingdom network analyses & SARS-CoV-2

To identify potential microbe-microbe and microbe-host interactions that could have an effect on outcome, we used a multi-scale network analysis approach (Multiscale Embedded Gene co-Expression Network Analysis, MEGENA)[39]. We first used the relative abundance from the RNAseq data to capture co-expressing taxa in the metatranscriptome network neighborhood of SARS-CoV-2 (SARS2-NWN). We examined five such network neighborhoods (constructed by including nodes with increasing distance 1 to 5 from SARS-CoV-2, i.e. neighborhood 1 to neighborhood 5) that were significantly enriched for taxa functionally active in the deceased group when compared with the 28-day MV group. Only the largest cluster, with 504 taxa, had significantly enriched taxa in both the deceased and in the 28-day MV outcome groups (Extended Data Fig. 10a) (FET P-value = 4.6e-45, 4.0 FE). Many of these taxa are among the top 50 most abundant microbes we had previously identified in the metatranscriptome dataset. Taxa present that are influenced by SARS-CoV-2 and significantly differentially enriched in the deceased group include bacteria such as *M. salivarium, Bifidobacterium breve*, and *Lactobacillus rhamnosus* (a gut commensal), that we had previously identified by differential expression analysis (Figure 3e), but also taxa such as *S. epidermis, Mycoplasma hominis* (urogenital bacteria), and the phage VB_PmiS-Isfahan (also referred to as Proteus virus Isfahan) that we had previously only picked up as being highly abundant but not necessarily differentially enriched in the deceased group. Most of the fungi, such as *C. albicans, C. glabrata* and *C. orthopsilosis* were enriched in the 28-day MV group. Interestingly, our earlier analysis of the metagenome (Extended Data Fig. 5b) had identified *C. glabrata* as being enriched in the deceased group with no enrichment in the metatranscriptome. This analysis indicates that some of these abundant taxa could be responding to SARS-CoV-2 disruption in a similar manner, or indirectly interacting functionally.

We further investigated the association of the network neighborhood with host network modules using the host transcriptome data to identify groups of host genes that are co-expressed in response to SARS-CoV-2 disruption. The 3 host modules with the most significant correlations to SARS2-NWN are M175, M277 and M718. M277 is the parent module of M718, and both are enriched with genes related to respiratory electron transport, while M175 is enriched for IFN-γ signaling (Extended Data Fig. 10b). Module M175 is positively correlated with the SARS2-NWN (ρ = 0.32, P-value = 2.1e-3). While there was no collective enrichment of the module by differentially expressed genes (DEGs) in the deceased vs 28-days MV, there was for >28-days vs 28-days MV (FET P-value = 0.030, 4.5 FE). This module includes well-known antiviral IFN stimulated genes (ISGs), such as *IRF7* and *OASL*.

**Metatranscriptome and host transcriptome signatures can predict mortality**

We evaluated the strength of the metatranscriptomic, metagenomic and host transcriptomic profiles to predict mortality in this cohort of critically ill COVID-19 patients. To this end, we identified features in each of these datasets and constructed risk scores that best predicted mortality. Figure 5a shows that the metatranscriptome data, alone or combined with the other two datasets, was most predictive of mortality. Importantly, the predictive power (as estimated by the area under the curve) of the metatranscriptome data was improved by excluding probable contaminants and worsened when SARS-CoV-2 was removed from the modeling. The selected features we used to construct the metatranscriptome, metagenome and host transcriptome risk scores are reported in Supplementary Table 8. Using the means of the scores, we classified all subjects into high risk and low risk groups for mortality. Figure 5b shows Kaplan-Meier survival curve comparisons evaluating the predictive power of risk score stratification based on metatranscriptome, metagenome and host transcriptome data. Combining risk scores from different datasets showed an optimal identification of mortality when metatranscriptome and host transcriptome were considered (Figure 5c). We then used the gene signature found as being the most predictive of mortality to conduct IPA analyses (Supplementary Table 9). Among the upstream regulators, mortality was associated with predicted activation of interferon alpha while chemotaxis and infection by RNA virus were predicted as activated in diseases and functions. These observations may be inconsistent with the current suggestion that, based on systemic levels, early interferon responses are associated with poor outcome in COVID19.[40,41] Others have suggested that a robust interferon response may lead to a hyperinflammatory state that could be detrimental in the disease process, justifying the use of Janus kinase inhibitor inhibitors in patients with COVID-19.[42] Studies comparing transcriptomic signatures in BAL of patients with severe COVID-19 and controls have shown activation of type 1 interferons.[43] While further longitudinal data will be needed to clarify the role of interferon signaling on the disease, the data presented here suggest that combining microbial and host signatures could help understand the increase risk for mortality in critically ill COVID-19 patients. Overall, these data highlight the importance of SARS-CoV-2 abundance in the lower airways as a predictor for mortality, and the significant contribution of the host cell transcriptome, which reflects the lower airway cell response to infection.

## Discussion

The samples used in this investigation were obtained during the first surge of cases of COVID-19 in New York City, and management reflected clinical practices at that time. Among the differences with current therapeutic approaches in COVID-19 patients, corticosteroids and remdesivir, two medications that likely affect the lower airway microbial landscape, were rarely used during the first surge. Other medications, such as antibiotics and anti-inflammatory drugs could affect our findings and we therefore considered them as potential confounders. However, the use of these medications was not found to be associated with clinical outcome. Of note, although our institutions were responding in "surge mode", both the Long Island and Manhattan campuses did not suffer from shortages in medical staff, supplies, or equipment and the decision to start mechanical ventilation did not differ from the standard of care. The cross-sectional study design precluded evaluation of the

temporal dynamics of the microbial community or the host immune response in this cohort, which could provide important insights into the pathogenesis of this disease. Performing repeated bronchoscopies without a clinical indication would be challenging in these patients and other less invasive methods might need to be considered to study the lower airways at earlier timepoints and serially over time in patients with respiratory failure. It is important to note that there were no statistically significant differences in the timing of sample collection across the three outcome groups. Evaluation of microbial signals at earlier timepoints in the disease process might also be important to identify changes occurring prior to broad spectrum antimicrobials use. Also, the presented data from lower airway samples are restricted to those subjects for whom bronchoscopy was performed as part of their clinical care. Thus, the culture independent data is biased towards patients that while critically ill with COVID-19 are not representative of the extremes in the spectrum of disease severity. Investigations focusing on early sample collection time points may be warranted to include subjects on mechanical ventilation with early mortality or early successful discontinuation of mechanical ventilation.

In summary, we present here the first evaluation of the lower airway microbiome using a metagenomic and metatranscriptomic approach, along with host immune profiling in critically ill patients with COVID-19 requiring invasive mechanical ventilation. The RNA metatranscriptome analysis showed an association between the abundance of SARS-CoV-2 and mortality, consistent with the signal found when viral load was assessed by targeted rRT-PCR. These viral signatures correlated with lower anti-SARS-CoV-2 Spike IgG and host transcriptomic signatures in the lower airways associated with poor outcome. Importantly, both through culture and NGS data, we did not find evidence for an association between untreated infections with secondary respiratory pathogens and mortality. Together, these data suggest that active lower airway SARS-CoV-2 replication and poor SARS-CoV-2-specific antibody responses are the main drivers of increased mortality in COVID-19 patients requiring mechanical ventilation. The potential role of oral commensals such as *Mycoplasma salivarium* need to be explored further. It is possible that *M. salivarium* can impact key immune cells and has recently been reported at a high prevalence in patients with ventilator-acquired pneumonia[44]. Critically, our finding that SARS-CoV-2 evades and/or derails effective innate/adaptive immune responses indicates that therapies aiming to control viral replication or induce a targeted antiviral immune response may be the most promising approach for hospitalized patients with SARS-CoV-2 infection requiring invasive mechanical ventilation.

## Methods

### Subjects

Enrolled subjects were 18 years or older, admitted to the intensive care units (ICUs) at NYU Langone Health from March 10th to May 10th, 2020 with a nasal swab confirmed diagnosis of SARS-CoV-2 infection by reverse transcriptase polymerase chain reaction (RT-PCR) assay and respiratory failure requiring invasive mechanical ventilation (see Table 1 for subject demographics). Research samples were obtained during clinically indicated bronchoscopies performed for airway clearance or for percutaneous tracheostomy placement

with verbal informed consent from legal authorized representative due to infection control measures that limited the presence of close contacts. All patients or their legal representative agreed to participate via our NYU IRB approved protocol (IRB # s16-00122/01598). Signed consent was then obtained from patients upon recovery. For those that remained incapacitated, signed consent was obtained from legally authorized representative. All analyses were then performed in de-identified data. Comprehensive demographic and clinical data were collected. We also collected longitudinal data on clinical laboratory culture results and treatment. Extended Data Fig 1 shows the distribution of subjects and sampling strategy used for this study. The study protocol was approved by the Institutional Review Board of New York University.

## Lower airway bronchoscopic sampling procedure

Both background and supraglottic (buccal) samples were obtained prior to the procedure, as previously described[23]. The background samples were obtained by passing sterile saline through the suctioning channel of the bronchoscope prior to the procedure. Bronchoalveolar lavage (BAL) samples were obtained from one lung segment as per discretion of the treating physician as clinically indicated. Samples were then transferred to a BSL3 laboratory for processing. Once there, 2 mL of whole BAL was stored in a tube prefilled with 2 mL of Zymo Research's DNA/RNA Shield™ (R1100-250, https://www.zymoresearch.com/pages/covid-19-efforts) for RNA/DNA preservation and virus inactivation. In addition, background control samples (saline passed through the bronchoscope prior to bronchoscopy) and supraglottic aspirates were stored in the same RNA/DNA shield.

## Viral load detection targeting the N gene

SARS-CoV-2 viral load was measured by quantitative real-time reverse transcription polymerase chain reaction (rRT-PCR) targeting the SARS-CoV-2 nucleocapsid (N) gene and an additional primer/probe set to detect the human RNase P gene (RP). Assays were performed using Thermo Fisher Scientific (Waltham, MA) TaqPath 1-Step RT-qPCR Master Mix, CG (catalog number A15299) on the Applied Biosystems (Foster City, CA) 7500 Fast Dx RealTime PCR Instrument. Using the positive controls provided by the CDC, which are normalized to 1000 copies/mL, we converted the different Ct positive to copies/mL. This was done using the DDCT method, applying the formula: Power [2, (CT (sample, N1 gene) – CT (PC, N1 gene)] – [CT (sample, RP gene) – CT (PC, RP gene)]*1000.

## SARS-CoV-2 viral viability through measurement of subgenomic transcripts

Viral subgenomic mRNA (sgRNA) is transcribed in infected cells and is not packaged into virions. Thus, presence of sgRNA is indicative of active infection of a mammalian cell in samples. We therefore measure sgRNA in all BAL samples obtained targeting the E gene as previously described.[19,20] Briefly, five μl RNA was used in a one-step real-time RT-PCR assay to sgRNA (forward primer 5'- CGATCTCTTGTAGATCTGTTCTC-3'; reverse primer 5'- ATATTGCAGCAGTACGCACACA-3'; probe 5'-FAMACACTAGCCATCCTTACTGCGCTTCG-ZEN-IBHQ-3') and using the Quantifast Probe RT-PCR kit (Qiagen) according to instructions of the manufacturer. In each run, standard dilutions of counted RNA standards were run in parallel to calculate copy numbers in the samples.

**Bacterial Load assessment—**We measured bacterial load in background, BAL and supraglottic samples using a QX200 Droplet Digital PCR System (BioRad, Hercules, CA). For this, primers were 5'-GCAGGCCTAACACATGCAAGTC-3' (63F) and 5'-CTGCTGCCTCCCGTAGGAGT-3' (355R). Cycling conditions included: 1 cycle at 95°C for 5 minutes, 40 cycles at 95°C for 15 seconds and 60°C for 1 minute, 1 cycle at 4°C for 5 minutes, and 1 cycle at 90°C for 5 minutes all at a ramp rate of 2°C/second. PCR cycling was performed on the BioRad C1000 Touch Thermal Cycler and droplets were quantified using the Bio-Rad Quantisoft software. Each sample was run in duplicate.

### DNA/RNA isolation, library preparation and sequencing

DNA and RNA were isolated in parallel using zymoBIOMICS™ DNA/RNA Miniprep Kit (Cat: R2002) as per manufacturer's instructions. DNA was then used for whole genome shotgun (WGS) sequencing using it as input into the NexteraXT library preparation kit following the manufacturer's protocol. Libraries were purified using the Agencourt AMPure XP beads (Beckman Coulter, Inc.) to remove fragments below 200 bp. The purified libraries were quantified using the Qubit dsDNA High Sensitivity Assay kit (Invitrogen) and the average fragment length for each library was determined using a High Sensitivity D1000 ScreenTape Assay (Agilent). Samples were added in an equimolar manner to form two sequencing pools. The sequencing pools were quantified using the KAPA Library Quantification Kit for Illumina platforms. The pools were then sequenced on the Illumina Novaseq 6000 in one single run. For RNA sequencing, RNA quantity and integrity were tested with a BioAnalyzer 2100 (Agilent). Among bronchoscope control (BKG) samples, only 5 yielded RNA with sufficient quality and quantity to undergo library preparation and sequencing. Further, in order to ensure sufficient depth on these background samples we used an equimolar strategy to pool the background samples based on the concentrations of each individual library. Of note, the same 5 BKG samples were selected to undergo WGS sequencing and we used the same pooling strategy. The automated Nugen Ovation Trio Low Input RNA method was used for library prep with 3ng total RNA input of each sample. After 6 amplification cycles, samples were sequenced using 2x Novaseq 6000 S4 200 cycle Flowcells using PE100 sequencing.

### Microbial community characterization using whole genome shotgun sequencing (WGS) and RNA metatranscriptome

For all metagenomic and metatranscriptomic reads, Trimmomatic v0.36[45], with leading and trailing values set to 3 and minimum length set to 36, was used to remove adaptor sequences. All rRNA reads were then removed from the metatranscriptomic reads using SortMeRNA v4.2.0[46] with default settings. Metagenomic and filtered metatranscriptomic reads were mapped to the human genome using Bowtie2 v2.3.4.1[47] with default settings and all mapping reads were excluded from subsequent microbiome, mycobiome, and virome metagenomic and metatranscriptomic analysis. Technical replicates for each biological sample were pooled together for subsequent analyses. Taxonomic profiles for all metagenomic and metatranscriptomic samples were generated using Kraken v2.0.7[48] and Bracken v2.5 [https://doi.org/10.7717/peerj-cs.104] run with default settings. The database used for quantifying taxonomic profiles was generated using a combined database containing human, bacterial, fungal, archaeal, and viral genomes downloaded

from NCBI RefSeq on January 8, 2021. Additionally, genomes for *Candida auris* (Genbank: GCA_003013715.2, GCA_008275145.1) and *Pneumocystis jirovecii* (Genbank: GCA_001477535.1) were manually added to the database. Supplementary Table 10 shows sequence depth and taxonomic richness per sample within sample types. Differentially abundant bacterial and viral taxa were identified for the BAL and UA samples groups individually using DESeq2 v1.28.1[49] with the three group clinical outcome meta-data readouts set as the sample groupings. Significantly differentially abundant taxa contained at a minimum an aggregate of 5 reads across samples and had an FDR <0.2[50,51]. The specificity of the top hits identified as being enriched by DESeq analysis were confirmed by mapping metagenomic and metatranscriptomics reads against whole genome references. In this manner we confirmed that sequence reads that were annotated to *M. salivarium*, *B. breve*, *L. rhamnosus*, *M. hominis*, and *S. oralis* (Fig. 3e) mapped along the length of the genomes and were not spurious matches (Supplementary Figure 3).

For functional microbial profiling, processed sequencing reads were further depleted of human-mapping reads by removing all reads classified as human by Kraken v2.0.7[48] using KrakenTools v0.1-alpha (https://github.com/jenniferlu717/KrakenTools). FMAP v0.15[52] was run on both the metagenomic and metatranscriptomic reads to profile the metabolic pathways present in each sample. FMAP_mapping.pl paired with diamond v0.9.24[53] and FMAP_quantification.pl were used with default settings to identify and quantify proteins in the Uniref90 database. Using DESeq2 v1.28.1[49], differentially expressed genes were identified for the BAL samples individually using the three group clinical outcome-metadata readouts for all genes that had an aggregate 5 reads across all samples.

Antibiotic resistance genes were quantified in all metagenome and metatranscriptome samples using Salmon v1.3.0[54] run with --keepDuplicates for indexing and --libtype A --allowDovetail --meta for quantification. Genes were filtered such that only genes that actively conferred antibiotic resistance were kept. To assess differentially expressed classes of antibiotic resistance genes, gene counts for individual antibiotic resistance genes were collapsed by their conferred antibiotic resistance.

Extended Data Fig. 1 shows a summary of depth achieved with the parallel WGS and metatranscriptome approach across sample types and the number of reads assigned to different microbial subfractions (bacteria, fungi, DNA viruses, RNA viruses and phages). Further analysis was also done to identify possible contaminants in the metatranscriptome and metagenome datasets. To this end, we compared the relative abundance of taxa between background bronchoscope control and BAL samples. Taxa with median relative abundance greater in background than in BAL were identified as probably contaminant and listed in Supplementary Table 4). None of the taxa identified as possible contaminants were removed from the analyzed data but are shown for comparison with signatures identified in the rest of the analyses.

### Anti-Spike SARS-CoV-2 antibody profiling in BAL

BAL samples were heat-treated at 56°C for one hour, and centrifuged at $14000 \times g$ for 5 min. The supernatant was collected and diluted 50-fold in PBST containing 1% skim milk. The diluted samples were incubated at room temperature (R.T.) for 30 min

with QBeads DevScreen: SAv (Streptavidin) (Sartorius 90792) that had been loaded with biotinylated Spike, biotinylated RBD or biotin (negative control) in wells of a 96 well HTS filter plate (MSHVN4550). As positive controls, we used CR3022 antibody, that recognizes SARS-CoV-2 Spike and RBD, in human IgG, IgA and IgM formats (Absolute Antibody; dilutions 1:1120, 1:1300 and 1:258, respectively). After washing the beads, bound antibodies were labeled with anti IgG-DyLight488, anti IgA-PE and anti IgM-PECy7, and the fluorescence intensities were measured in Intellicyt IQue3 (Sartorius). The acquired data [median fluorescence intensity (MFI)] were normalized using the MFI values of the CR3022 antibodies to compensate for variations across plates. Extended Data Fig 8 shows that the levels of these antibodies were higher in BAL samples of patients with SARS-CoV-2 than in BAL samples from 10 uninfected healthy smokers recruited for research bronchoscopy. Details of method development and validation will be described elsewhere (Koide et al. in preparation).

**SARS-CoV-2 preparation and neutralization assay**

icSARS-CoV-2-mNG (isolate USA/WA/1/2020, obtained from the UTMB World Reference Center for Emerging Viruses and Arboviruses) was amplified once in Vero E6 cells (P1 from the original stock). Briefly, 90–95% confluent T175 flask (Thomas Scientific) of Vero E6 ($1\times10^7$ cells) was inoculated with 50 μL of icSARS-CoV-2-mNG in 5 mL of infection media (DMEM, 2% FBS, 1% NEAA, and 10 mM HEPES) for 1 hour. After 1 hour, 20 mL of infection media was added to the inoculum and cells were incubated 72 hours at 37 °C and 5% $CO_2$. After 72 hours, the supernatant was collected and the monolayer was frozen and thawed once. Both supernatant and cellular fractions were combined, centrifuged for 5 min at $500 \times g$, and filtered using a 0.22 μm Steriflip (Millipore). Viral titers were determined by plaque assay in Vero E6 cells. In brief, 220,000 Vero E6 cells/well were seeded in a 24 well plate, 24 hours before inoculation. Ten-fold dilutions of the virus in DMEM (Corning) were added to the Vero E6 monolayers for 1 hour at 37 °C. Following incubation, cells were overlaid with 0.8% agarose in DMEM containing 2% FBS (Atlanta biologicals) and incubated at 37 °C for 72 h. The cells were fixed with 10% formalin, the agarose plug removed, and plaques visualized by crystal violet staining. All procedures including icSARS-CoV-2-mNG virus were performed using Biosafety Level 3 laboratory conditions.

For SARS-CoV-2 neutralization assays, Vero E6 cells (30,000 cells/well) were seeded in a 96 well plate 24 h before infection. Two-fold serial dilutions of BAL lysates were mixed with mixed 1:1 (vol/vol) with SARS-CoV-2 mNG virus (multiplicity of infection, MOI 0.5), and incubated for 1 h at 37 °C. After incubation, 100 μL of the mixtures of the antibody and SARS-CoV-2 mNG were added to the Vero E6 monolayers, and cells were incubated at 37°C. After 20 h, cells were fixed with 4 % formaldehyde (Electron Microscopy Sciences) at room temperature for 1 h. After fixation, cells were washed twice with PBS and permeabilized with 0.25% triton-100, stained with DAPI (Thermo), and quantified on a CellInsight CX7 High-content microscope (Thermo) using a cut-off for three standard deviations from negative to be scored as an infected cell.

### Transcriptome of BAL cells

RNA-Seq was performed on bronchial epithelial cells obtained by airway brushing, as described[55–57], using the Hi-seq/Illumina platform at the NYU Langone Genomic Technology Center (data available at Sequence Read Archive: # PRJNA592149). KEGG[58,59] annotation was summarized at levels 1 to 3. Genes with an FDR-corrected adjusted p-value <0.25 were considered significantly differentiated, unless otherwise specified. Pathway analysis using differentially regulated genes (FDR<0.25) was done using Ingenuity Pathway Analysis, RRID:SCR_0- at least 1 count per million in at least two samples were retained. For digital cytometry with CIBERSORTx, a signature matrix derived from single-cell transcriptome of BAL cells collected from patients with COVID-19[38] was first generated with the "Create Signature Matrix" module in the CIBERSORTx online tool. A maximum of 10 cells per cell type per patient were initially sampled from the original data and 20 cells per cell type were then used to build the single-cell reference with the default parameters. Then the "Impute Cell Fractions" module was used to estimate the absolute cell fraction score of different cell types in bulk transcriptomes using the single-cell signatures with "S-mode" batch correction and 100 permutations in the absolute mode. Bulk transcriptomes with a significant deconvolution p-value ( 0.05) were retained. For xCell cell type signature enrichment analysis, the enrichment scores were inferred with built-in signature of cell types detected in the BAL samples as reported previously 38. The two-tailed Wilcoxon rank sum test with Benjamini-Hochberg correction were computed between groups of samples for comparison.

### Microbial and Host predictive modeling

Cox proportional hazards model was used for investigating the association between the time to death and the relative abundance of each taxon quantified using metatranscriptomic and metagenomic data separately. We first performed the univariate screening test to identify significant features associated with the time to death using the Cox proportion hazards regression model for the relative abundance of taxa from the RNA and DNA data, and log-transformed count of host transcriptome data, respectively. Within each type of data, given the p-value cutoff, the features with a p-value less than the cutoff were selected and integrated as a sub-community. For the RNA and DNA data, the alpha diversity (Shannon index) was calculated for each sample on the selected sub-community and the negative of the value was defined as the microbial risk score, because high alpha diversity indicates low risk of death. For the host transcriptome data, the log-transformed total count of all selected candidate transcriptome for each sample was defined as the risk score, since most selected candidate transcriptomes increased the risk of death. The leave-one-out cross-validation (LOOCV) was used for the predictions. The p value cutoff was set at the value which produces the largest AUC (area under the receiver operating characteristic curve) in predicting the death/survival status using the risk score we constructed over these features. The additive model was used to integrate when more than one scores are used for the prediction.

## Multiscale and co-expression network analyses

Raw counts from the human transcriptome were normalized and converted to log2-counts per million using limma[60]/voom[61] (v3.44.1 with R v4.0.0) with standard parameters. Microbiome abundance information was converted to relative abundance. Low abundance taxa were removed based on average abundance across all samples to yield a minimum of 1000 taxa for each metatranscriptome dataset. All datasets were batch adjusted. Differentially expressed genes (DEGs) and differentially abundant taxa were called using the DESeq2 package[49] (v1.28.1), based on the negative binomial (i.e. Gamma-Poisson) distribution. According to the recommendation by the authors, we used non-normalized data (i.e. raw gene counts and abundance data), as DESeq2 internally corrects data and performs normalization steps. For this purpose, raw microbiome abundance data were converted to DESeq2 dds objects using the phyloseq R library (V1.32.0). Contrasts are based on outcome groups ( 28 days MV, > 28 days MV or death). Differentially expressed genes and differentially abundant tax with FDR of 0.2 or less are considered significant.

Multiscale Embedded Gene Co-Expression Network Analysis (MEGENA) 39 was performed to identify host modules of highly co-expressed genes in SARS-CoV-2 infection. The MEGENA workflow comprises four major steps: 1) Fast Planar Filtered Network construction (FPFNC), 2) Multiscale Clustering Analysis (MCA), 3) Multiscale Hub Analysis (MHA), 4) and Cluster-Trait Association Analysis (CTA). The total relevance of each module to SARS-CoV-2 infection was calculated by using the Product of Rank method with the combined enrichment of the differentially expressed gene (DEG) signatures as implemented: $G_j = \Pi_i g_{ji}$, where, $g_{ji}$ is the relevance of a consensus $j$ to a signature $i$; and $g_{ji}$ is defined as $(max_j(r_{ji}) + 1 - r_{ji})/\Sigma_j r_{ji}$ where $r_{ji}$ is the ranking order of the significance level of the overlap between the module $j$ and the signature.

To functionally annotate gene signatures and gene modules derived from the host transcriptome data, we performed an enrichment analysis of the established pathways and signatures—including the gene ontology (GO) categories and MSigDB. The hub genes in each subnetwork were identified using the adopted Fisher's inverse Chi-square approach in MEGENA; Bonferroni-corrected p-values smaller than 0.05 were set as the threshold to identify significant hubs. The correlation between modules, modules and clinical traits as well as modules and individual taxa were performed using Spearman correlation. Other correlation measures, such as Pearson correlation or the Maximal Information Coefficient (MIC)[62] proved to be inferior for this task. Categorical trait data was converted to numerical values as suitable.

## Statistics & Reproducibility

Specific statistical analysis is described in detail above. For association with discrete factors, we used non-parametric tests (Mann-Whitney or Kruskal-Wallis ANOVA). We used the ade4 package in R to construct Principal Coordinate Analysis (PCoA) based on weighted UniFrac distances[63,64]. To cluster microbiome communities into exclusive 'metacommunities' we used a Dirichlet Multinomial Mixture Model using the R package DirichletMultinomial[65,66]. No statistical method was used to predetermine sample size. No data were excluded from

the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment.

### Data availability

All sequencing data used for this analysis are available in NCBI's Sequence Read Archive under project numbers PRJNA688510 and PRJNA687506 (RNA and DNA sequencing, respectively).

### Code availability

Codes used for the analyses presented in the current manuscript are available at https://github.com/segalmicrobiomelab/SARS_CoV2.

## Extended Data
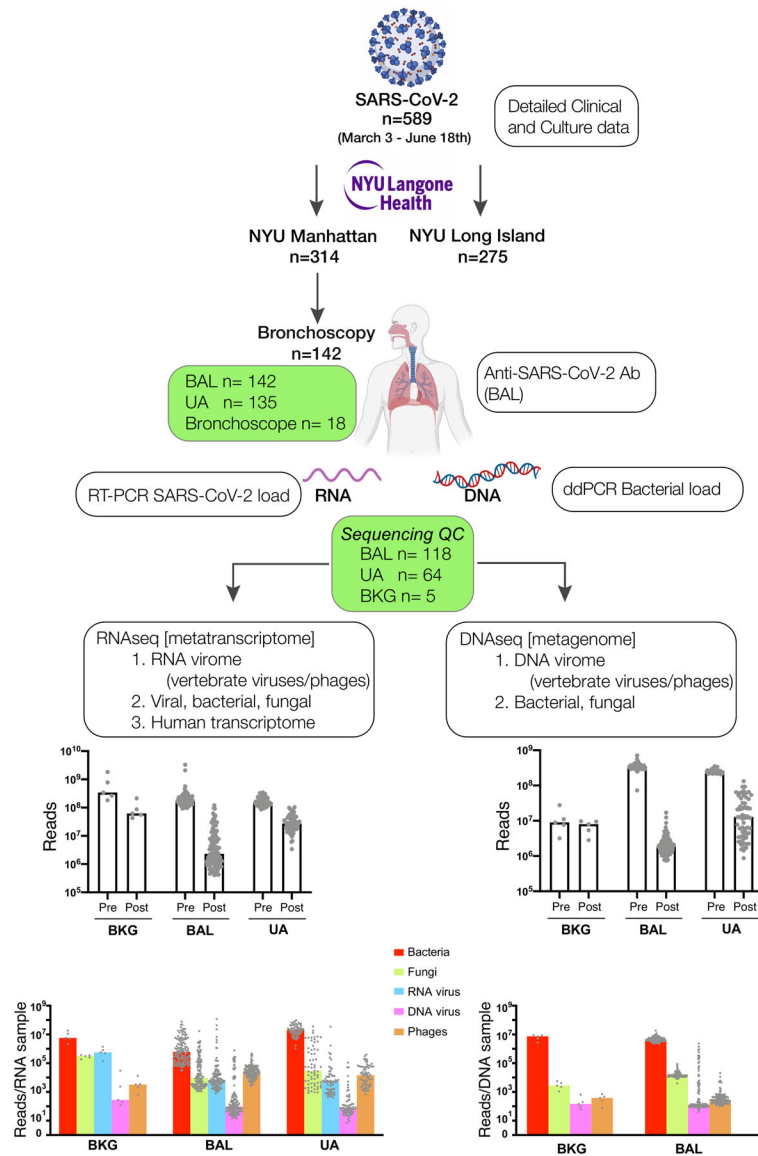


**Extended Data Fig. 1. Description of patient cohort, samples obtained, analyses performed and sequencing depth.**

Schematic representation of the measurements obtained in this cohort of 589 critically ill COVID-19 patients. Top barplots summarize median number of reads obtained per sample of 142 independent subjects prior to filtering (Pre) and after filtering (Post). Bottom barplots show the median number of reads annotated to different microbial kingdoms in each sample.

*SARS-CoV-2, Human with Lungs, RNA and DNA images were created with

BioRender.com

**NYU Langone Health is the official brand from NYU Langone Health.
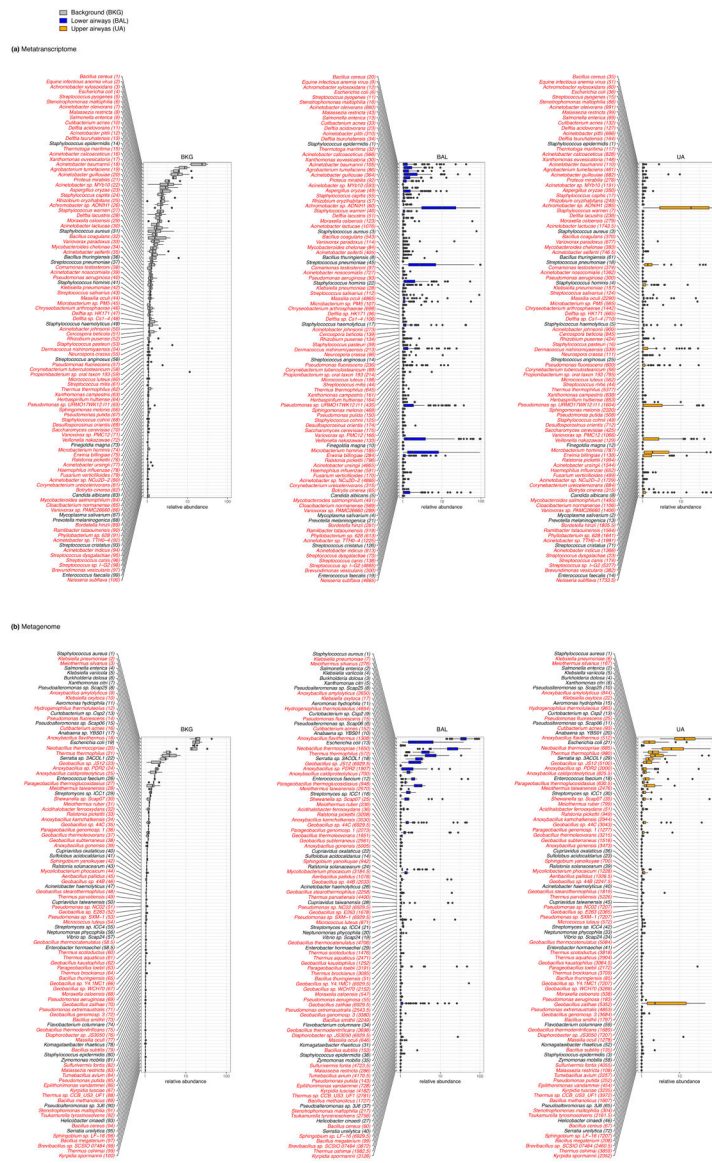
**Extended Data Fig. 2. Identification of top taxa found in background samples as compared with BAL and upper airway samples.**
Boxplots showing the relative abundance values in log10 relative abundance of taxa ranked ordered based on dominance of Background bronchoscope control samples and compared to abundances in BAL and Upper Airway samples within metatranscriptome **(a)** and metagenome **(b)** data. Red labels indicate taxa where relative abundance is higher in background samples than in BAL and therefore considered possible contaminant.
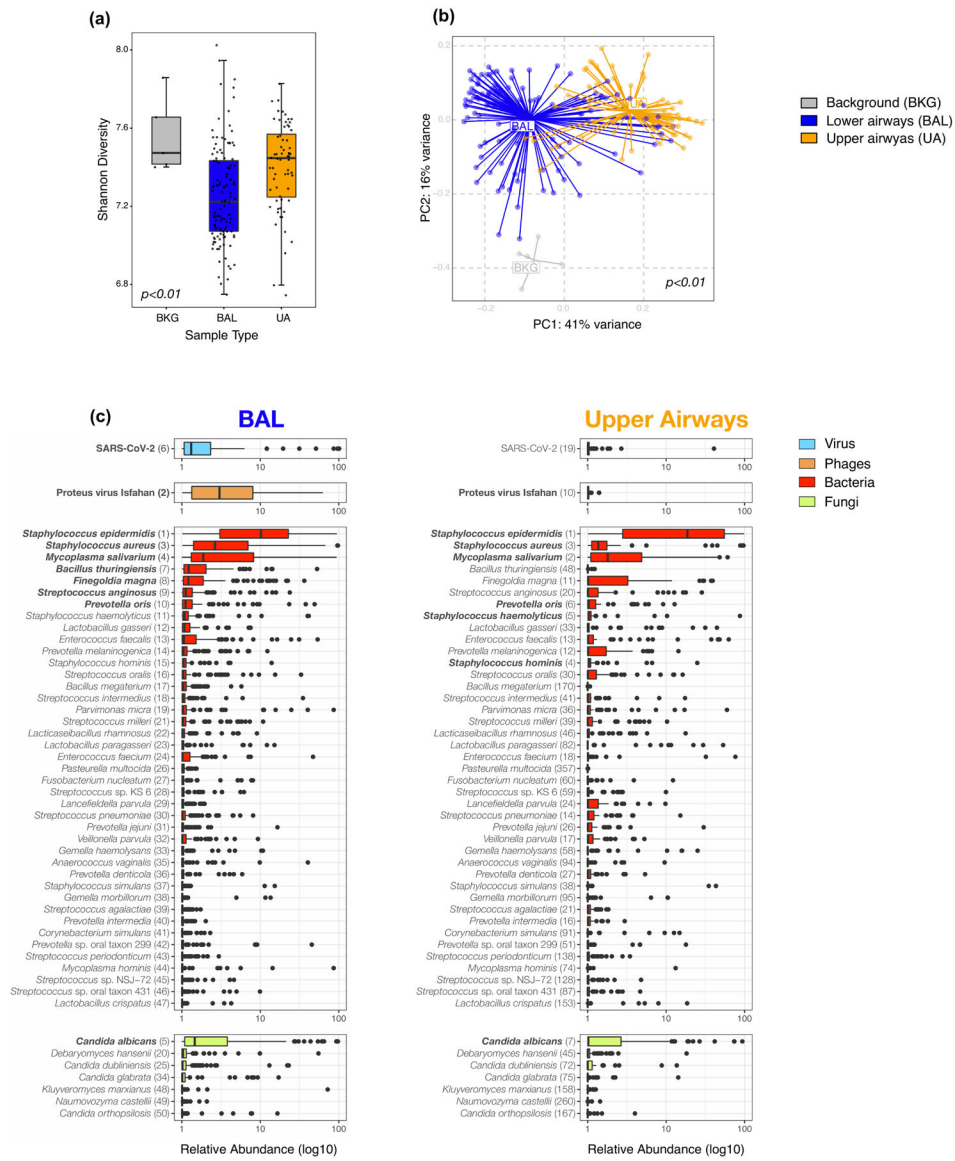
**Extended Data Fig. 3. Topographical analyses of metatranscriptome data.**
**a**) Comparison of alpha diversity (Shannon Index, each dot denotes the Shannon diversity
of a sample while the box center depicts median, box interquartile range with median at
the center and whiskers represent maximum and minimum value) and **b**) beta diversity
(Bray Curtis Dissimilarity index, across 5 background negative controls (bronchoscope), 118
bronchoalveolar lavage (BAL) and 64 upper airway (UA) samples (Kruskal-Wallis p value
=0.0006 and PERMANOVA p-values = 0.001, without multiple comparisons, respectively).
**(c)** Boxplots showing the relative abundance values in log10 across all metatranscriptome
samples for the 118 BAL and 64 Upper Airway samples. The 50 taxa with the highest
relative abundance values in the BAL metatranscriptome data are displayed; the top 10
in the BAL are highlighted in bold. Each column consists of four plots displaying in
decreasing order of abundance the top RNA vertebrate viruses, DNA phages, bacteria, and
fungi identified (from top to bottom). Numbers in parentheses next to the taxa labels display

the ranking in relative abundance for either the BAL or UA metatranscriptome samples, respectively. Each dot denotes the relative abundance of a taxa per sample while the box center depicts median, box inter-quartile range with median at the center and whiskers represent maximum and minimum value.
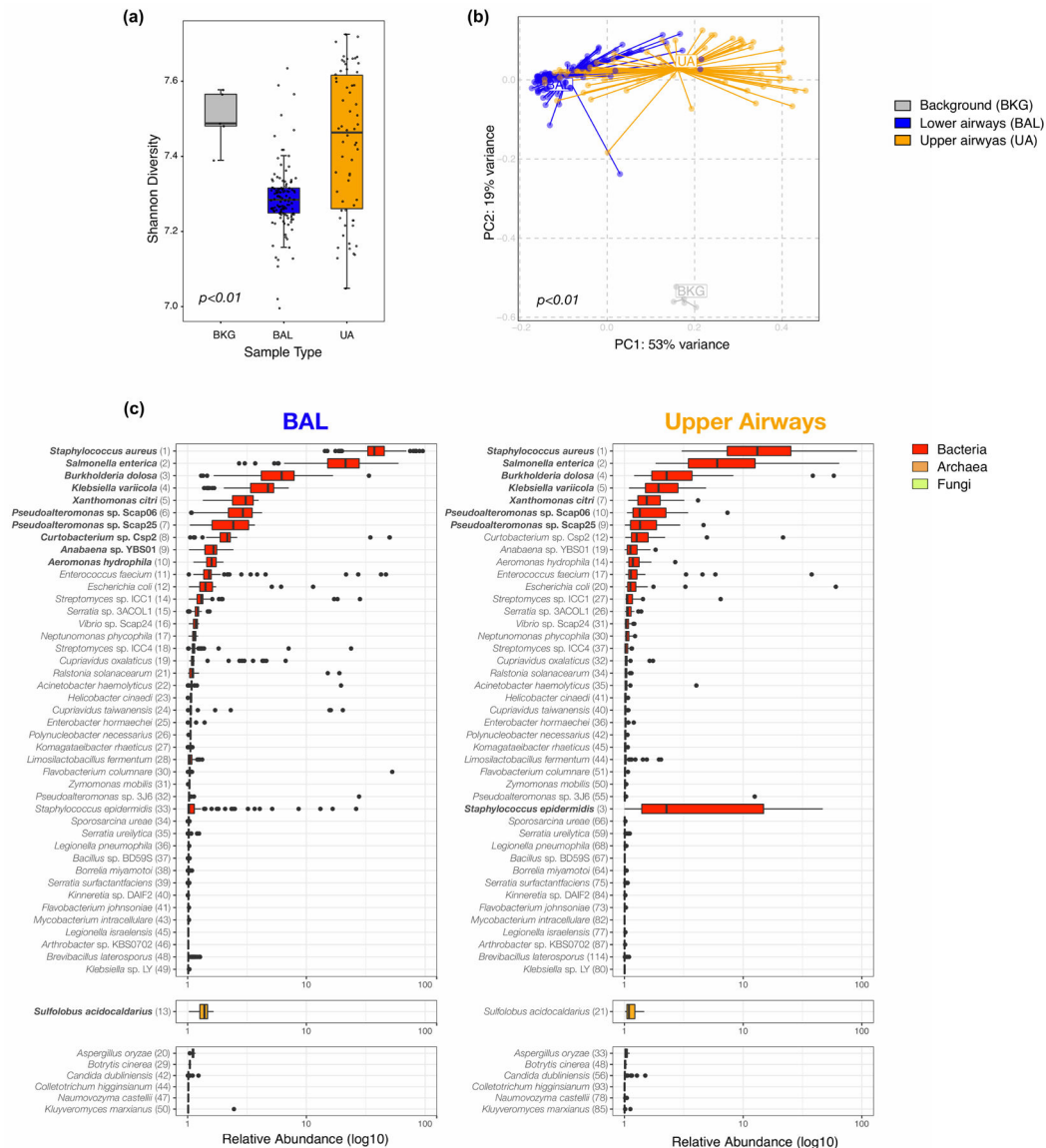


**Extended Data Fig. 4. Topographical analyses of metagenome data**
Comparison of **a**) alpha diversity (Shannon Index, each dot denotes the Shannon diversity of a sample while the box center depicts median, box interquartile range with median at the center and whiskers represent maximum and minimum value) and **b**) beta diversity (Bray Curtis Dissimilarity index, across 5 background negative controls (bronchoscope), 118 bronchoalveolar lavage (BAL) and 64 upper airway (UA) samples (Kruskal-Wallis p-value = 0.00000000000000022 and PERMANOVA p-value= 0.001, without multiple comparisons, respectively). (c) Boxplots showing the relative abundance values in log10

across all metagenome samples for the 118 BAL and 64 Upper Airway samples. The 50 taxa with the highest relative abundance values in the BAL metagenome are displayed; the top 10 in the BAL are highlighted in bold. Each column consists of two plots displaying the most abundant bacteria and fungi identified. Numbers in parentheses next to the taxa labels displays its ranking in relative abundance for either the BAL or UA metagenome samples, respectively. Each dot denotes the relative abundance of a taxa per sample while the box center depicts median, box inter-quartile range with median at the center and whiskers represent maximum and minimum value.
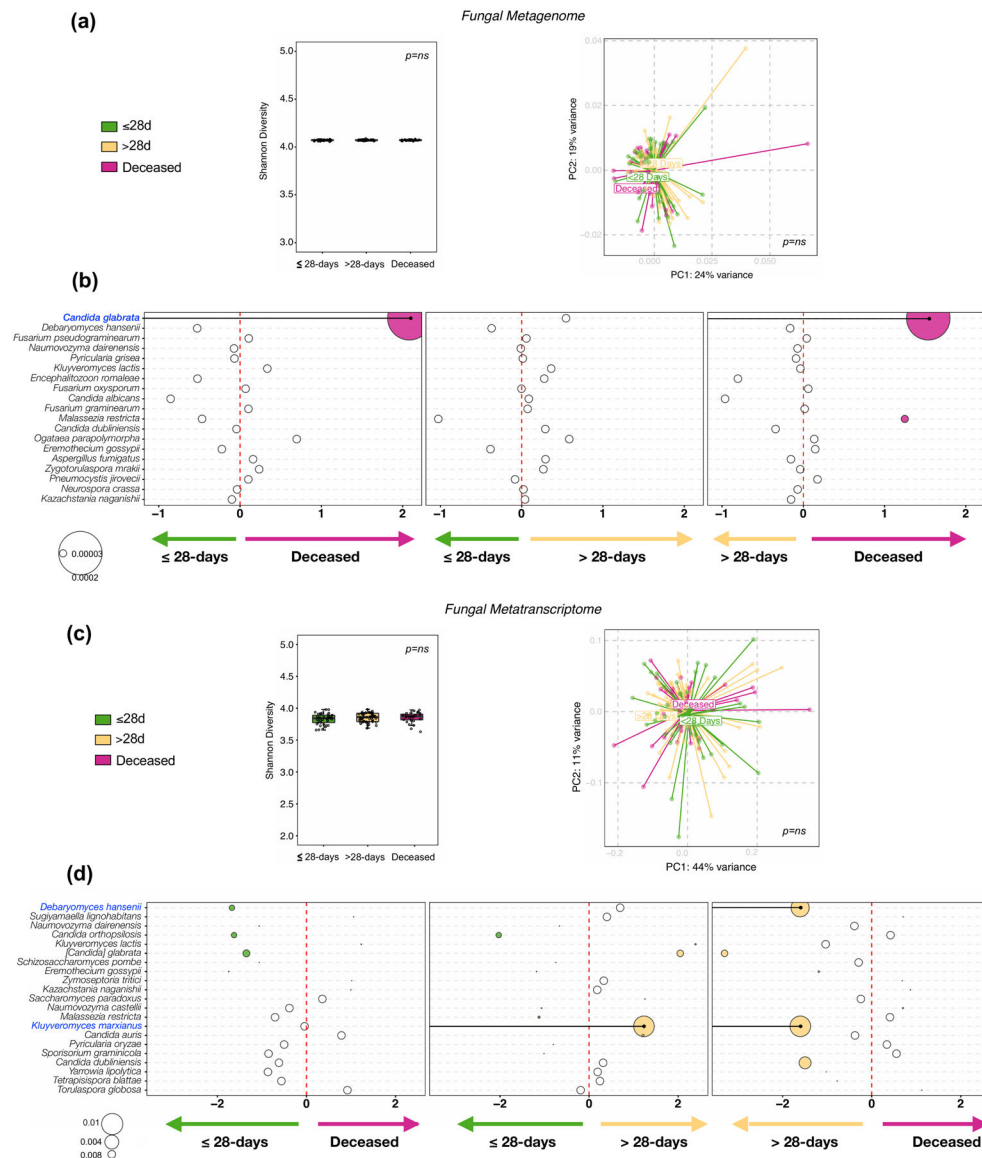


**Extended Data Fig. 5. Evaluation of associations between the lower airway mycobiome and clinical outcome.**

Fungal taxonomic data was subtracted from metagenome and metatranscriptome data from 5 background negative controls (bronchoscope), 118 bronchoalveolar lavage (BAL) and 64 upper airway (UA) samples.. **(a)** Comparisons between the three clinical outcome groups

was performed for α diversity (Shannon Index, each dot denotes the Shannon diversity of a sample while the box center depicts median, box inter-quartile range with median at the center and whiskers represent maximum and minimum value, left panel), β diversity (based on Bray Curtis Dissimilarity Index, right panel); Kruskal-Wallis p-value and PERMANOVA p-value respectively; on metagenome data. **(b)** Bubble plot showing DESeq results of fungi enriched in each clinical outcome comparisons based on metagenome data (bubble size based on median relative abundance for those found to be statistically significant). **(c)** Comparisons between the three clinical outcome groups was performed for α diversity (Shannon Index, each dot denotes the Shannon diversity of a sample while the box center depicts median, box inter-quartile range with median at the center and whiskers represent maximum and minimum value, left panel), β diversity (based on Bray Curtis Dissimilarity Index, right panel); Kruskal-Wallis p-value and PERMANOVA p-value respectively; on metatranscriptome data. **(d)** Bubble plot showing DESeq results of fungi enriched in each clinical outcome comparisons based on metatranscriptome data (bubble size based on median relative abundance for those found to be statistically significant).
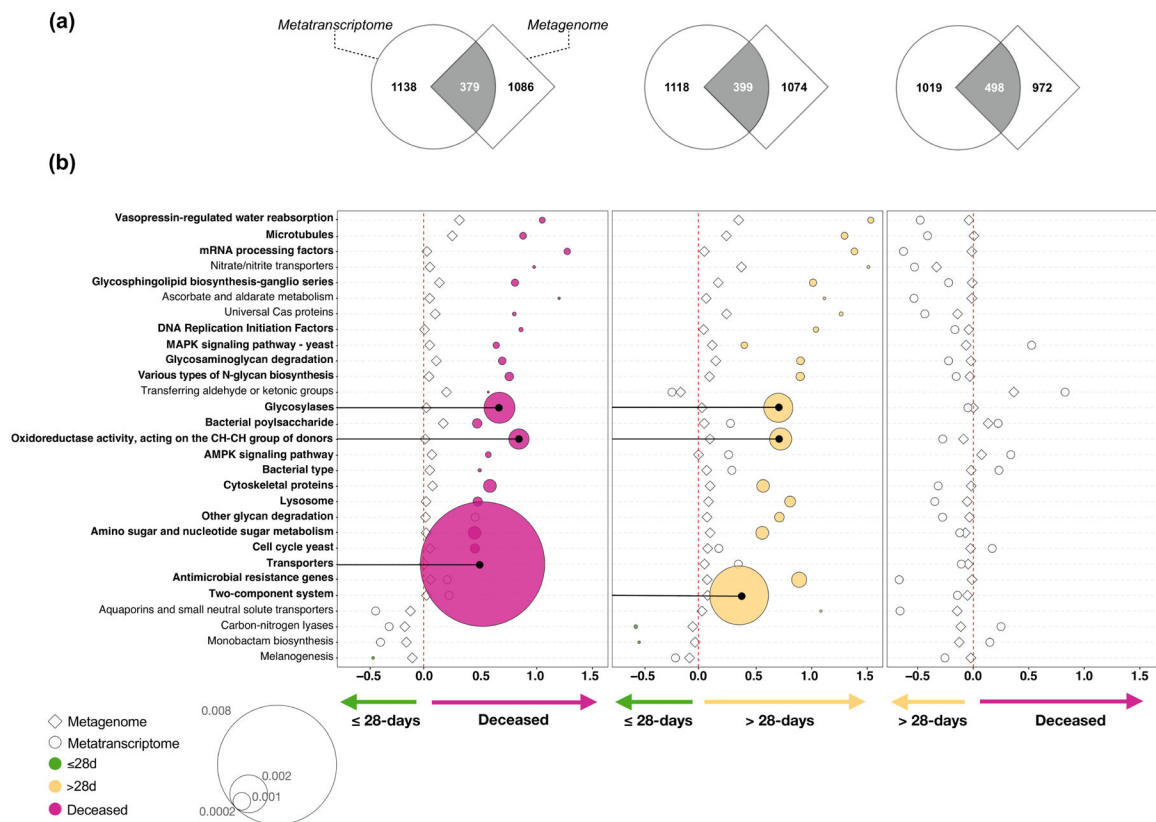


**Extended Data Fig. 6. Functional microbial compositional analyses.**
KOs were summarized to associated pathways and differential expression was calculated based on DESeq2 analysis. **(a)** Gene Set Enrichment Analysis (GSEA) was used to compare the functional signatures identified in BAL metagenome and metatranscriptome as distinctly enriched for comparisons between clinical outcome groups. **(b)** Bubble plot showing DESeq results of microbial functions found concordantly differentially enriched between

clinical outcome groups (bubble size based on median relative abundance for those found statistically significant).
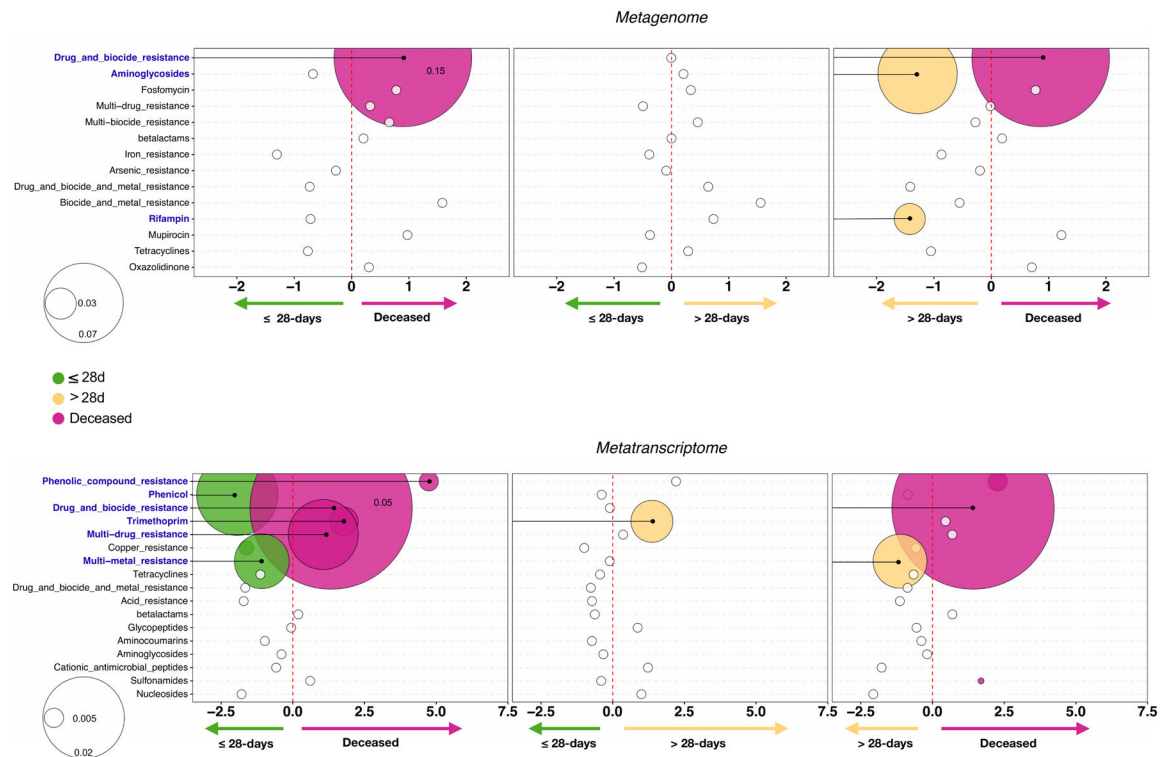


**Extended Data Fig. 7. Evaluation of associations between the lower airway antibiotic resistance genes and clinical outcome.**

Bubble plot showing DESeq results of summarized categories of antibiotic resistant microbial genes taken from MEGARes for the metagenome (top) and metatranscriptome (bottom) data sets for each clinical outcome comparison (bubble size based on median relative abundance for those found to be statistically significant). Colored bubbles indicate significantly enriched antibiotic resistance groups.
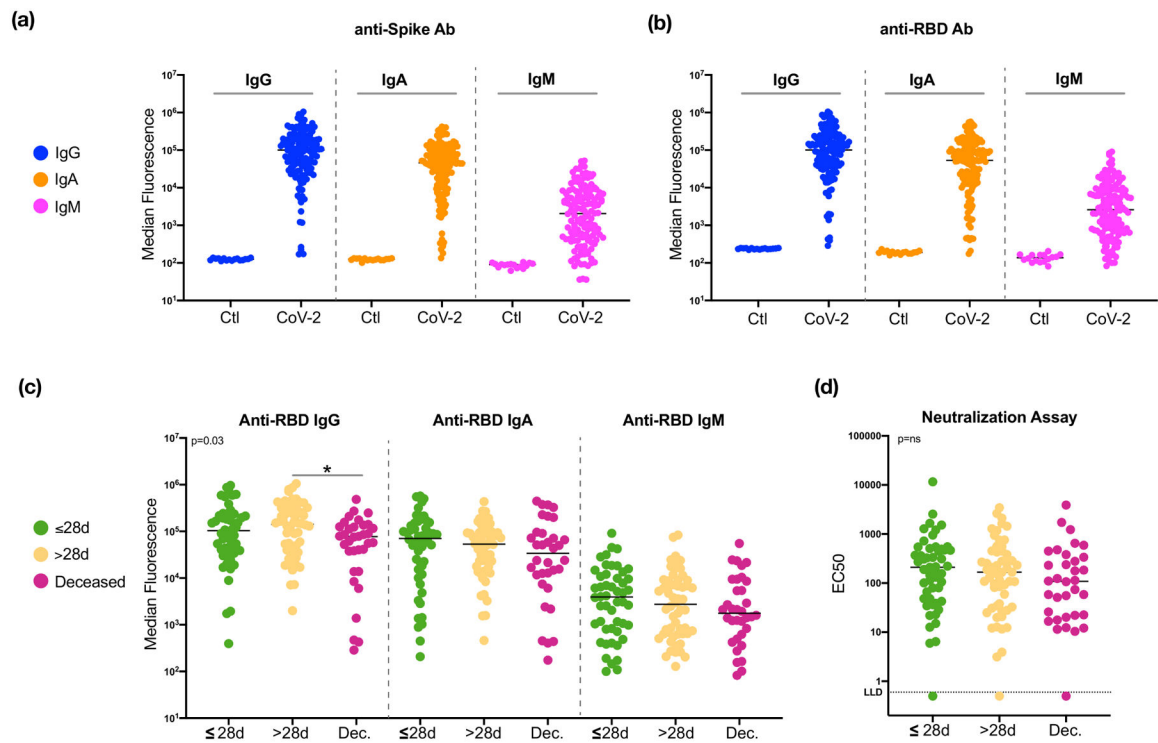
**Extended Data Fig. 8. Measurement of anti-SARS-CoV-2 Immunoglobulin levels and neutralization activity.**

Levels of anti-SARS-CoV-2 Spike **(a)** and anti-SARS-CoV-2 receptor binding domain (RBD, **b**) antibodies in from 20 non-SARS-CoV-2 infected smoker controls and 142 severely ill COVID-19 intubated patients. Note that the signals for different isotypes cannot be compared because they are detected with different reagents. **(c)** Comparisons of anti-SARS-CoV-2 RBD antibody levels in 142 BAL samples across subjects in different clinical outcome groups (*= Two-sided Mann–Whitney U p<0.05)**. (d)** Neutralizing activity in BAL samples across subjects in different clinical outcome groups.

**(a)**



**(b)** **(c)** **(d)**



**Extended Data Fig. 9. Evaluation for associations between the lower airway host transcriptome and clinical outcome**

**(a)** PCoA (based on Bray Curtis Dissimilarity Index, PERMANOVA p-value) comparing the three clinical outcome groups. **(b, c, d)** Volcano plots comparing lower airway host transcriptome between the three clinical outcome groups.

**Extended Data Fig. 10. Multi-scale cross-kingdom and co-expression networks.**
The neighborhood 5 cross-kingdom metatranscriptome network centered around SARS-CoV-2 is shown. Nodes refer to taxa, edges denote co-abundance after MEGENA. The size of the nodes indicates abundance. Taxa with large nodes are highly abundant. Node shapes are according to the legend and refer to different microbial kingdoms. The differential abundance of taxa in log2(fold change) between the deceased group and the 28-day MV groups is shown by node color - red nodes are taxa abundant in the deceased group compared to the 28-day MV group, blue colored nodes denote the opposite. **(b)** Modules M175 and M718 of the host transcriptome are shown. The node size refers to the absolute gene expression value. Nodes with wide node border refer to key regulators/hub genes (see Methods). The differential gene expression of taxa in log2(fold change) between the deceased group and the 28-day MV groups is shown by node color - red nodes are

up-regulated in the deceased group compared to the 28-day MV group, blue colored nodes denote the opposite.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Authors

Imran Sulaiman[1,2], Matthew Chung[3], Luis Angel[1,2], Jun-Chieh J. Tsay[1,2,4], Benjamin G. Wu[1,2,4], Stephen T. Yeung[5], Kelsey Krolikowski[1,2], Yonghua Li[1,2], Ralf Duerr[5], Rosemary Schluger[1,2], Sara A. Thannickal[5], Akiko Koide[2,6], Samaan Rafeq[1,2], Clea Barnett[1,2], Radu Postelnicu[1,2], Chang Wang[7], Stephanie Banakis[3], Lizzette Pérez-Pérez[8], Guomiao Shen[9], George Jour[9], Peter Meyn[10], Joseph Carpenito[1,2], Xiuxiu Liu[1,10], Kun Ji[1,11], Destiny Collazo[1,2], Anthony Labarbiera[1,2], Nancy Amoroso[1,2], Shari Brosnahan[1,2], Vikramjit Mukherjee[1,2], David Kaufman[1,2], Jan Bakker[1,2], Anthony Lubinsky[1,2], Deepak Pradhan[1,2], Daniel H. Sterman[1,2], Michael Weiden[1,2], Adriana Hegu[9,12], Laura Evans[13], Timothy M. Uyeki[14], Jose C. Clemente[15], Emmie de Wit[8], Ann Marie Schmidt[16], Bo Shopsin[17], Ludovic Desvignes[5], Chan Wang[18], Huilin Li[18], Bin Zhang[19], Christian V. Forst[19], Shohei Koide[6,20], Kenneth A. Stapleford[5], Kamal M. Khanna[5,6], Elodie Ghedin[3,7,*], Leopoldo N. Segal[1,2,6,*]

## Affiliations

[1]Division of Pulmonary and Critical Care Medicine, New York University Grossman School of Medicine, NYU Langone Health, New York, NY

[2]Department of Medicine, New York University Grossman School of Medicine, NYU Langone Health, New York, NY

[3]Systems Genomics Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

[4]Division of Pulmonary and Critical Care Medicine, VA New York Harbor Healthcare System, NY

[5]Department of Microbiology, New York University Grossman School of Medicine, NYU Langone Health, New York, NY.

[6]Laura and Isaac Perlmutter Cancer Center, New York University School of Medicine, NYU Langone Health, New York, NY.

[7]Center for Genomics & Systems Biology, Department of Biology, New York University New York, NY

[8]Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rocky Mountain Laboratories, Hamilton, MT.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

[9]Department of Pathology, New York University Grossman School of Medicine, NYU Langone Health, New York, NY.

[10]Division of Pediatrics, Longhua Hospital Affiliated to Shanghai University of Chinese Medicine, CN

[11]Dongfang Hospital affiliated to Beijing University of Traditional Chinese Medicine, Beijing, China

[12]NYU Langone Genome Technology Center, New York University Grossman School of Medicine, NYU Langone Health, New York, NY

[13]Pulmonary, Critical Care and Sleep Medicine, University of Washington, Seattle, USA

[14]Influenza Division, Centers for Disease Control and Prevention, Atlanta, GA

[15]Department of Genetics and Genomic Sciences and Immunology Institute, Icahn School of Medicine at Mount Sinai, NY.

[16]Diabetes Research Program, Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, New York University Grossman School of Medicine, NYU Langone Health, New York, NY

[17]Division of Infectious Diseases, Department of Medicine, New York University School of Medicine, NYU Langone Health, New York, NY

[18]Department of Population Health, New York University School of Medicine, NYU Langone Health, New York, NY

[19]Department of Genetics and Genomic Sciences, Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY

[20]Department of Biochemistry and Molecular Pharmacology, New York University Grossman School of Medicine, NYU Langone Health, New York, NY and Perlmutter Cancer Center, New York, NY 10016

## Acknowledgements:

## Reference:

1. The, L. Emerging understandings of 2019-nCoV. Lancet395, 311 (2020). [PubMed: 31986259]

2. WHO coronavirus disease (COVID-19) dashboard. Geneva: World Health Organization, Available online: https://covid19.who.int/ (2020).

3. Rabaan AA, et al.SARS-CoV-2, SARS-CoV, and MERS-COV: A comparative overview. Infez Med28, 174–184 (2020). [PubMed: 32275259]

4. Cao XCOVID-19: immunopathology and its implications for therapy. Nat Rev Immunol20, 269–270 (2020). [PubMed: 32273594]

5. Morens DM & Fauci AS The 1918 influenza pandemic: insights for the 21st century. J Infect Dis 195, 1018–1028 (2007). [PubMed: 17330793]

6. Shieh WJ, et al.2009 pandemic influenza A (H1N1): pathology and pathogenesis of 100 fatal cases in the United States. Am J Pathol177, 166–175 (2010). [PubMed: 20508031]

7. Dickson RP, et al.Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. Nature microbiology1, 16113 (2016).

8. Kitsios GD, et al.Respiratory Tract Dysbiosis Is Associated with Worse Outcomes in Mechanically Ventilated Patients. Am J Respir Crit Care Med202, 1666–1677 (2020). [PubMed: 32717152]

9. Dickson RP, et al.Lung Microbiota Predict Clinical Outcomes in Critically Ill Patients. Am J Respir Crit Care Med201, 555–563 (2020). [PubMed: 31973575]

10. Zuo T, et al.Alterations in Fecal Fungal Microbiome of Patients With COVID-19 During Time of Hospitalization until Discharge. Gastroenterology159, 1302–1310 e1305 (2020). [PubMed: 32598884]

11. Chen L, et al.RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerg Microbes Infect9, 313–319 (2020). [PubMed: 32020836]

12. Shen Z, et al.Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients With Coronavirus Disease 2019. Clin Infect Dis71, 713–720 (2020). [PubMed: 32129843]

13. Kelleni MTSARS CoV-2 viral load might not be the right predictor of COVID-19 mortality. J Infect (2020).

14. Fajnzylber J, et al.SARS-CoV-2 viral load is associated with increased disease severity and mortality. Nat Commun11, 5493 (2020). [PubMed: 33127906]

15. Bitker L, et al.Protracted viral shedding and viral load are associated with ICU mortality in Covid-19 patients with acute respiratory failure. Ann Intensive Care10, 167 (2020). [PubMed: 33301059]

16. Magleby R, et al.Impact of SARS-CoV-2 Viral Load on Risk of Intubation and Mortality Among Hospitalized Patients with Coronavirus Disease 2019. Clin Infect Dis (2020).

17. Westblade LF, et al.SARS-CoV-2 Viral Load Predicts Mortality in Patients with and without Cancer Who Are Hospitalized with COVID-19. Cancer Cell38, 661–671 e662 (2020). [PubMed: 32997958]

18. Pujadas E, et al.SARS-CoV-2 viral load predicts COVID-19 mortality. Lancet Respir Med8, e70 (2020). [PubMed: 32771081]

19. Wolfel R, et al.Virological assessment of hospitalized patients with COVID-2019. Nature581, 465–469 (2020). [PubMed: 32235945]

20. Kim D, et al.The Architecture of SARS-CoV-2 Transcriptome. Cell181, 914–921 e910 (2020). [PubMed: 32330414]

21. Speranza E, et al.Single-cell RNA sequencing reveals SARS-CoV-2 infection dynamics in lungs of African green monkeys. Science translational medicine13(2021).

Author Manuscript    Author Manuscript    Author Manuscript    Author Manuscript

22. Yazdi M, Bouzari M & Ghaemi EA Genomic analyses of a novel bacteriophage (VB_PmiS-Isfahan) within Siphoviridae family infecting Proteus mirabilis. Genomics 111, 1283–1291 (2019). [PubMed: 30149052]

23. Tsay JJ, et al.Airway Microbiota Is Associated with Upregulation of the PI3K Pathway in Lung Cancer. Am J Respir Crit Care Med198, 1188–1198 (2018). [PubMed: 29864375]

24. Tsay JJ, et al.Lower airway dysbiosis affects lung cancer progression. Cancer Discov (2020).

25. Sulaiman I, et al.Evaluation of the airway microbiome in nontuberculous mycobacteria disease. Eur Respir J52(2018).

26. Sterlin D, et al.IgA dominates the early neutralizing antibody response to SARS-CoV-2. Science translational medicine13(2021).

27. Wang Z, et al.Enhanced SARS-CoV-2 neutralization by dimeric IgA. Science translational medicine13(2021).

28. Klingler J, et al.Role of IgM and IgA Antibodies in the Neutralization of SARS-CoV-2. J Infect Dis (2020).

29. Budayeva HG, Rowland EA & Cristea IM Intricate Roles of Mammalian Sirtuins in Defense against Viral Pathogens. J Virol 90, 5–8 (2016). [PubMed: 26491165]

30. Dar HH, et al.Pseudomonas aeruginosa utilizes host polyunsaturated phosphatidylethanolamines to trigger theft-ferroptosis in bronchial epithelium. J Clin Invest128, 4639–4653 (2018). [PubMed: 30198910]

31. Stoyanovsky DA, et al.Iron catalysis of lipid peroxidation in ferroptosis: Regulated enzymatic or random free radical reaction?Free Radic Biol Med133, 153–161 (2019). [PubMed: 30217775]

32. Qiang Z, et al.Nrf2 and STAT3 Alleviates Ferroptosis-Mediated IIR-ALI by Regulating SLC7A11. Oxid Med Cell Longev2020, 5146982 (2020). [PubMed: 33014271]

33. Xu Y, Li X, Cheng Y, Yang M & Wang R Inhibition of ACSL4 attenuates ferroptotic damage after pulmonary ischemia-reperfusion. FASEB J 34, 16262–16275 (2020). [PubMed: 33070393]

34. Hallman M, Bry K, Hoppu K, Lappi M & Pohjavuori M Inositol supplementation in premature infants with respiratory distress syndrome. N Engl J Med 326, 1233–1239 (1992). [PubMed: 1560798]

35. Preuss S, et al.Inositol-trisphosphate reduces alveolar apoptosis and pulmonary edema in neonatal lung injury. Am J Respir Cell Mol Biol47, 158–169 (2012). [PubMed: 22403805]

36. Newman AM, et al.Determining cell type abundance and expression from bulk tissues with digital cytometry. Nature Biotechnology37, 773–782 (2019).

37. Aran D, Hu Z & Butte AJ xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biology 18, 220 (2017). [PubMed: 29141660]

38. Liao M, et al.Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nature Medicine26, 842–844 (2020).

39. Song WM & Zhang B Multiscale Embedded Gene Co-expression Network Analysis. PLoS computational biology 11, e1004574 (2015). [PubMed: 26618778]

40. Bastard P, et al.Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science370(2020).

41. Zhang Q, et al.Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science370(2020).

42. Kalil AC, et al.Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. N Engl J Med (2020).

43. Zhou Z, et al.Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. Cell host & microbe27, 883–890 e882 (2020). [PubMed: 32407669]

44. Nolan TJ, et al.Low-pathogenicity Mycoplasma spp. alter human monocyte and macrophage function and are highly prevalent among patients with ventilator-acquired pneumonia. Thorax71, 594–600 (2016). [PubMed: 27071419]

45. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]

46. Kopylova E, Noe L & Touzet H SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217 (2012). [PubMed: 23071270]

47. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359 (2012). [PubMed: 22388286]

48. Wood DE, Lu J & Langmead B Improved metagenomic analysis with Kraken 2. Genome biology 20, 257 (2019). [PubMed: 31779668]

49. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15, 550 (2014). [PubMed: 25516281]

50. Pavel AB, et al.Alterations in Bronchial Airway miRNA Expression for Lung Cancer Detection. Cancer Prev Res (Phila)10, 651–659 (2017). [PubMed: 28877936]

51. Seumois G, et al.Transcriptional Profiling of Th2 Cells Identifies Pathogenic Features Associated with Asthma. J Immunol197, 655–664 (2016). [PubMed: 27271570]

52. Kim J, Kim MS, Koh AY, Xie Y & Zhan X FMAP: Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies. BMC Bioinformatics 17, 420 (2016). [PubMed: 27724866]

53. Buchfink B, Xie C & Huson DH Fast and sensitive protein alignment using DIAMOND. Nat Methods 12, 59–60 (2015). [PubMed: 25402007]

54. Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14, 417–419 (2017). [PubMed: 28263959]

55. Mortazavi A, Williams BA, McCue K, Schaeffer L & Wold B Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5, 621–628 (2008). [PubMed: 18516045]

56. Wilhelm BT, et al.Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature453, 1239–1243 (2008). [PubMed: 18488015]

57. Sultan M, et al.A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science321, 956–960 (2008). [PubMed: 18599741]

58. Tanabe M & Kanehisa M Using the KEGG database resource. Curr Protoc Bioinformatics Chapter 1, Unit1 12 (2012).

59. Kanehisa M, Goto S, Sato Y, Furumichi M & Tanabe M KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40, D109–114 (2012). [PubMed: 22080510]

60. Law CW, et al.RNA-seq analysis is easy as 1–2-3 with limma, Glimma and edgeR. F1000Res5(2016).

61. Law CW, Chen Y, Shi W & Smyth GK voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol 15, R29 (2014). [PubMed: 24485249]

62. Reshef DN, et al.Detecting novel associations in large data sets. Science334, 1518–1524 (2011). [PubMed: 22174245]

63. Dray S.a.D., A.B.The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software22, 1–20 (2007).

64. Lozupone C, Lladser ME, Knights D, Stombaugh J & Knight R UniFrac: an effective distance metric for microbial community comparison. The ISME journal 5, 169–172 (2011). [PubMed: 20827291]

65. Holmes I, Harris K & Quince C Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One 7, e30126 (2012). [PubMed: 22319561]

66. Morgan MDirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data. (2017).
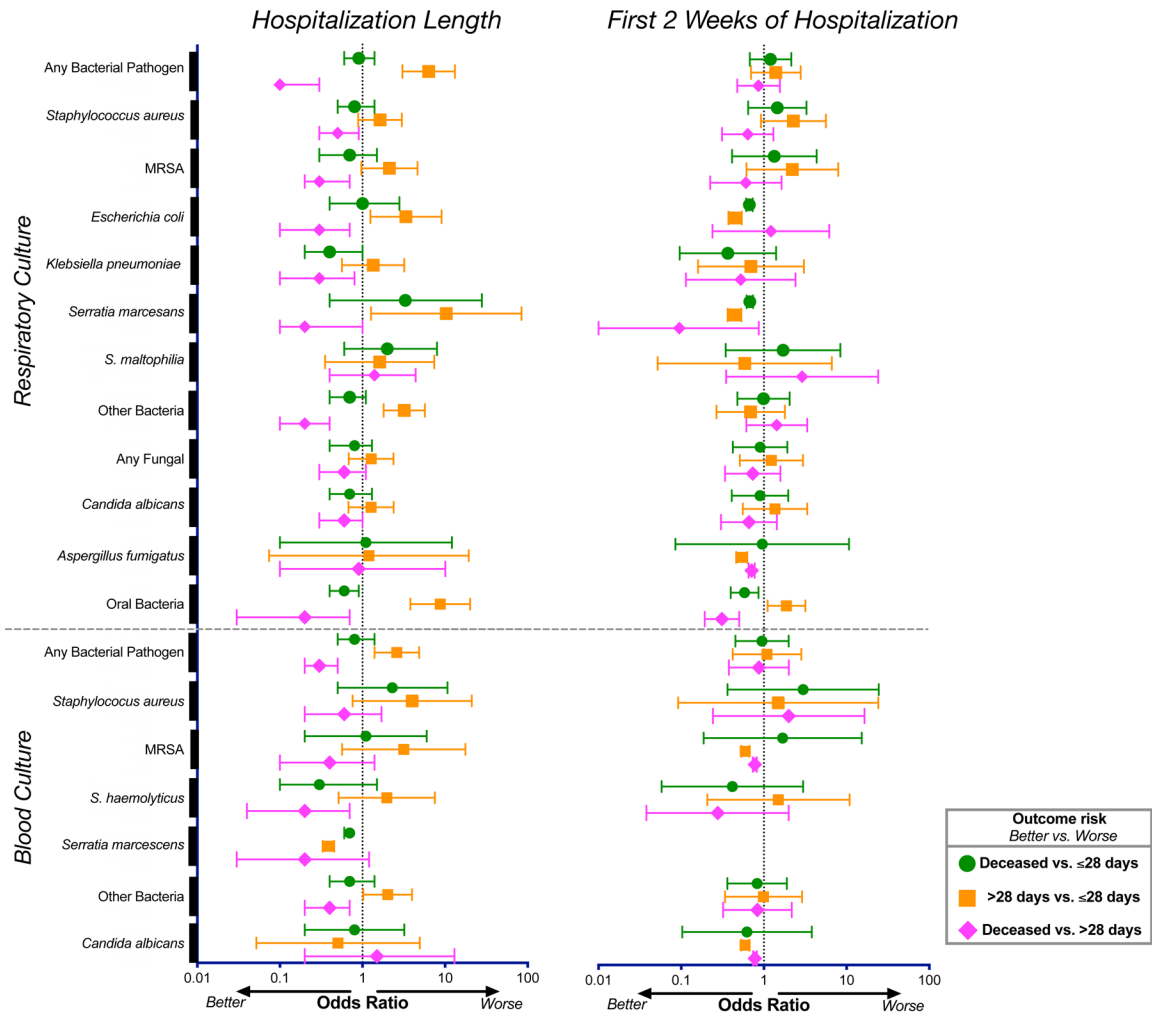
**Figure 1. Associations between culture positivity and clinical outcome.**
Odds ratios and corresponding 95% confidence intervals for rates of culture positivity for the whole cohort (n=589) during length of hospitalization (left) and during the first 2 weeks of hospitalization (right).
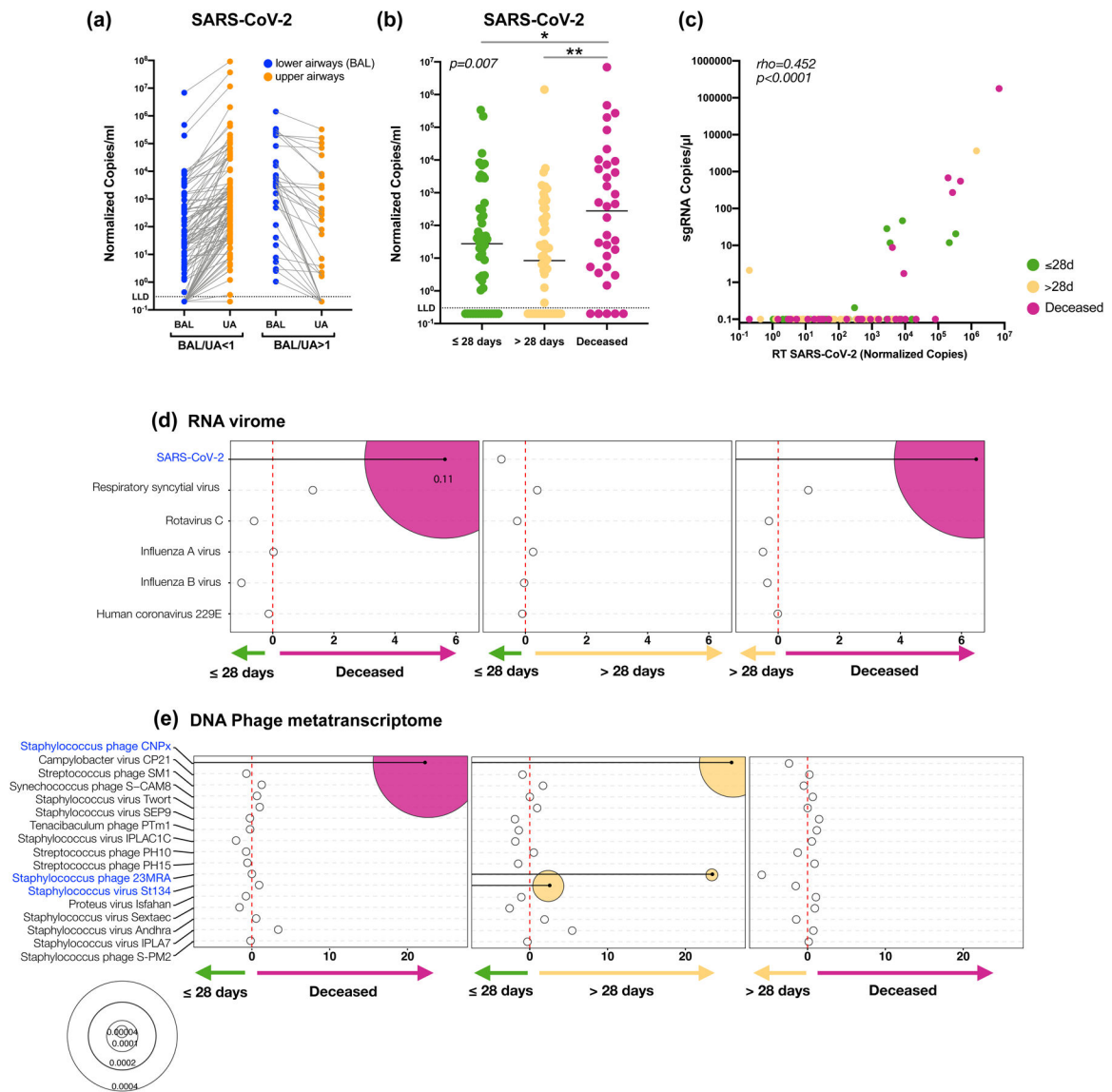
**Figure 2. SARS-CoV-2 viral load and virus metatranscriptome analyses.**
Copies of the SARS-CoV-2 N gene per ml, normalized by the Human RNase P gene, comparing paired upper and lower airway samples **(a)** and levels in BAL comparing clinical outcome groups in 142 subjects **(b**, *= Two-sided Mann–Whitney U p=0.038, **= Two-sided Mann–Whitney U p=0.001;). **(c)** PCoA analysis based on Bray Curtis Dissimilarity index of BAL Metatranscriptome data comparing clinical outcome (PERMANOVA p-value). Bubble plot showing DESeq results of RNA viruses **(d)** and expressed DNA phages **(e)** enriched in each clinical outcome comparisons (bubble size based on median relative abundance for those found to be statistically significant).
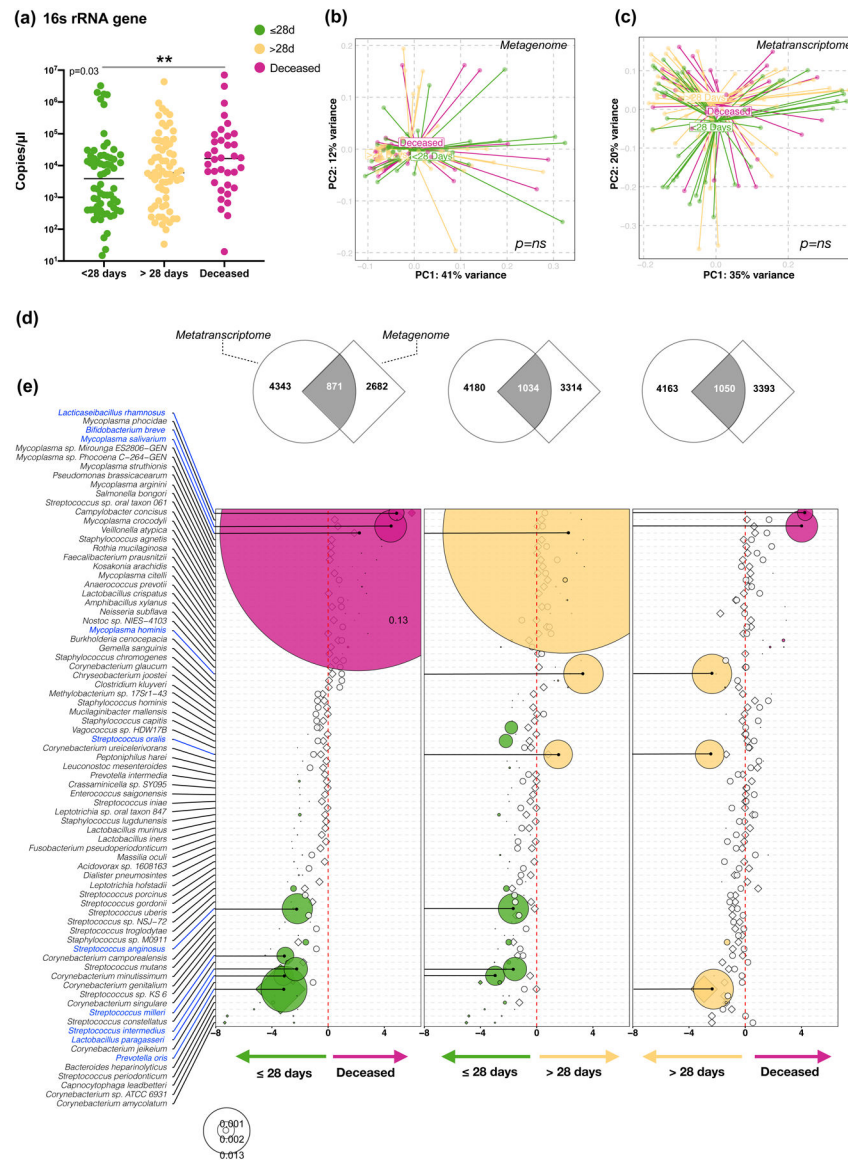
**Figure 3. Bacteria load and taxonomic compositional analyses.**
(a) Bacterial load measured by ddPCR targeting 16S rRNA gene in 142 subjects (**=
Two-sided Mann–Whitney U p<0.01). PCoA analysis based on Bray Curtis Dissimilarity
index of BAL (b) Metagenome (c) and Metatranscriptome data comparing clinical outcome
(Single variable PERMANOVA p-value). (d) Gene Set Enrichment Analysis (GSEA) was
used to compare the taxonomic signatures identified in BAL metagenome (diamonds) and
metatranscriptome (circles) as distinctly enriched for comparisons between clinical outcome
groups (differential enrichment performed based on DESeq2 analysis). (e) Bubble plot
showing DESeq results of bacteria found concordantly differentially enriched between
clinical outcome groups (bubble size based on median relative abundance for those found
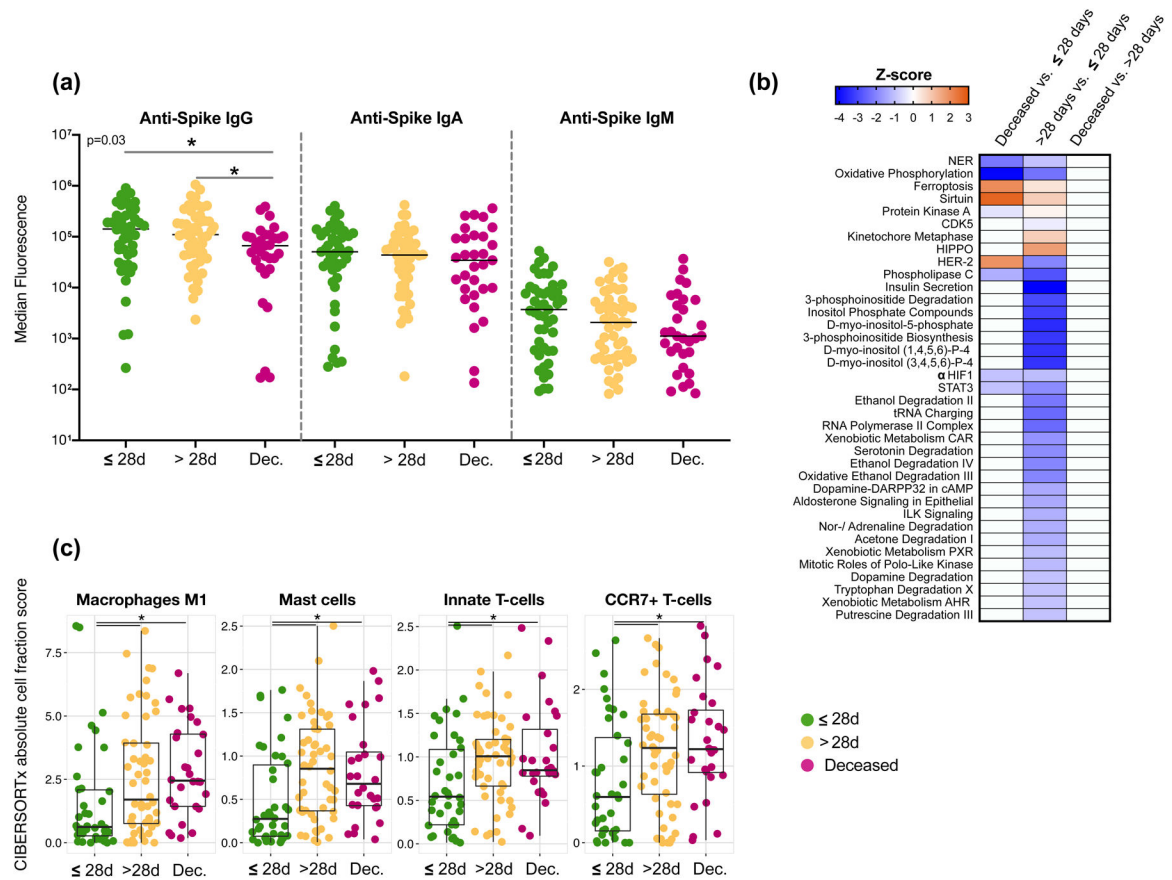statistically significant).

**Figure 4. Lower airway host immune profiling in severely ill COVID-19.**
(a) Levels of anti-SARS-CoV-2 Spike antibodies in BAL of 142 subjects (*= Two-sided
Mann–Whitney U; Deceased vs <28D p=0.0133; Deceased vs >28D p=0.0213). (b)
Heat-map of canonical pathway analysis based on Ingenuity Pathway Analysis (IPA,
RRID:SCR_008653) using the lower airway host transcriptome comparing clinical outcome
groups. Orange shows up-regulation of pathway, blue shows down-regulation of pathway.
(c) Cell type abundance quantification plots. Comparison of abundance of mast cells
and neutrophils among outcome groups in the BAL fluids of 118 critically ill patients
with COVID-19. Cell type abundance was estimated from the host transcriptome with
CIBERSORTx. Each dot denotes the quantification score of a sample while the box center
depicts median, box inter-quartile range with median at the center and whiskers represent
maximum and minimum (*=Two-sided Mann–Whitney U; Macrophages: M1 >28D vs
<28D p=0.017; Deceased vs <28D p=0.0014; Mast Cells: >28D vs <28D p=0.0013;
Deceased vs <28D p=0.0056; Innate T-Cells: >28D vs <28D p=0.0071; Deceased vs <28D
p=0.0068; CCR7+ T-cells: >28D vs <28D p=0.0192; Deceased vs <28D p=0.0068).
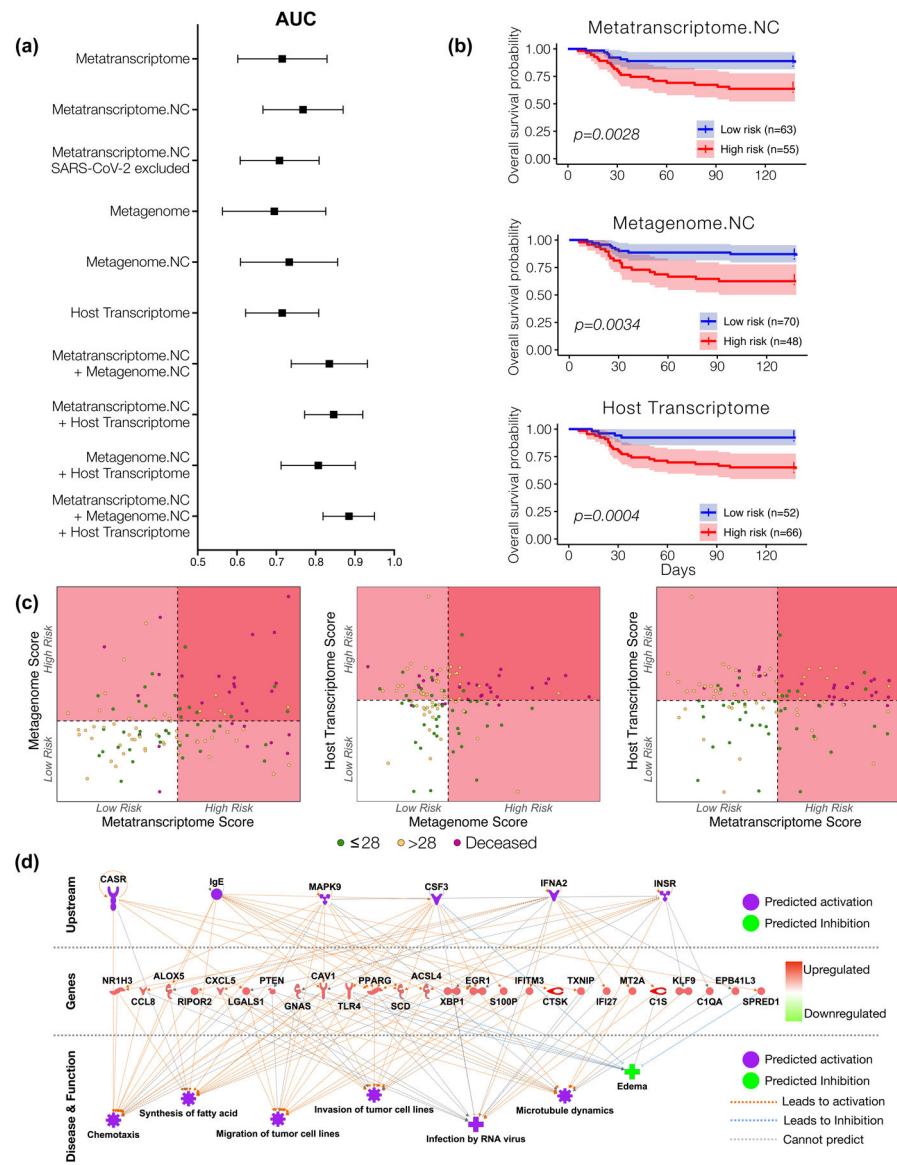
**Figure 5. Mortality predictive power of metatranscriptome, metagenome and host transcriptome.**
**(a)** Area under the curve median and 95% confidence interval for receiver operating characteristic curve analyses calculated from each sequencing datasets as predictor and mortality (death time from ICU admission) as outcome from 118 subjects. **(b)** Kaplan-meier survival analyses with 95% confidence interval based on a cutoff value estimated from features selected from each sequencing dataset. The "High risk" and "Low risk" groups is the mean of predicted risk scores in all samples (p value with Matel Cox). **(c)** Scatterplot among risk scores from metatranscriptome, metagenome, and host transcriptome. Dotted line denotes the mean of the risk scores across all subjects, which is also the threshold for dividing the samples into "High risk" and "Low risk" groups. **(d)** IPA analyses of host transcriptomic signatures identified as most predictive of mortality.

**Table 1.**

NYU Manhattan Bronchoscopy Comparison Cohort

| Variable | Outcomes | | | |
|---|---|---|---|---|
| | MV 28 days | MV >28 days | Deceased | *p-value* |
| **N** | 52(36.6) | 56(39.4) | 34(24) | |
| **Age** | 59[38–67] | 64[47–71] | 64[56–72] | 0.094 |
| **Sex (Male)** | 40(76.9) | 46(82.1) | 25(73.5) | 0.608 |
| **Race/Ethnicity** | | | | 0.231 |
| Caucasian | 25(48.1) | 28(50) | 14(41.2) | |
| Hispanic or Latino | 10(19.2) | 16(28.6) | 8(23.5) | |
| AAM | 1(1.9) | 4(7.1) | 6(17.6) | |
| Asian | 4(7.7) | 4(7.1) | 1(2.9) | |
| Other | 22(42.3) | 20(35.7) | 13(38.2) | |
| **BMI** | 29[25–32] | 26[23–29] | 29[25–33] | 0.094 |
| **Comorbidities** | | | | |
| Hyperlipidemia | 11(21.2) | 19(33.9) | 7(20.6) | 0.226 |
| Hypertension | 29(55.8) | 23(41.1) | 17(50) | 0.306 |
| CHF | 2(3.8) | 3(5.4) | 3(8.8) | 0.615 |
| CAD | 5(9.6) | 8(14.3) | 5(14.7) | 0.705 |
| Diabetes | 18(34.6) | 23(41.1) | 13(38.2) | 0.788 |
| Asthma | 1(1.9) | 0(0) | 1(2.9) | 0.478 |
| CKD | 4(7.7) | 3(5.4) | 8(23.5) | *0.017* [*#] |
| CVA | 3(5.8) | 10(17.9) | 13(38.2) | *0.001* [*#] |
| **Smoking Status** | | | | 0.846 |
| Ever | 10(19.2) | 13(23.2) | 8(23.5) | |
| Never | 42(80.8) | 43(76.8) | 26(76.5) | |
| **Bio-Markers** [¥] | | | | |
| IL-6 | 83[44–180] | 40[16–143] | 113[23–214] | 0.284 |
| Lymphocytes | 9[7–12] | 6[4–8] | 4[3–6] | *6.984E-11* [*#$] |
| WBC | 10.7[8.9–13.4] | 12.2[9.8–15.7] | 14.2[11–17.4] | *0.004* [*] |
| Ferritin | 1286[722–2513] | 1448[915–2352] | 1882[1001–2893] | 0.240 |
| CRP | 49[23–135] | 82[36–138] | 66[37–157] | 0.224 |
| D-Dimer | 2038[956–3592] | 2350[747–3399] | 2006[889–3035] | 0.903 |
| $PaO_2/FiO_2$ | 168[103–210] | 96[74–178] | 97[65–152] | *0.001* [*$] |
| **Treatment** | | | | |
| ECMO | 10(19.2) | 15(26.8) | 2(5.9) | *0.05* [#] |
| Dialysis | 5(9.6) | 15(26.8) | 11(32.4) | *0.023* [*$] |
| Steroids | 26(50) | 44(78.6) | 28(82.4) | *0.001* [*$] |
| Anticoagulation | 50(96.2) | 54(96.4) | 34(100) | 0.521 |

| | Outcomes | | | |
|---|---|---|---|---|
| **Variable** | **MV 28 days** | **MV >28 days** | **Deceased** | *p-value* |
| Hydroxychloroquine | 49(94.2) | 51(91.1) | 31(91.2) | 0.799 |
| Tocilizumab | 24(46.2) | 23(41.1) | 12(35.3) | 0.604 |
| **Antiviral** | 18(34.6) | 21(37.5) | 14(41.2) | 0.827 |
| Lopinavir/Ritonavir | 10(19.2) | 8(14.3) | 6(17.6) | 0.784 |
| Remdesivir | 5(9.6) | 8(14.3) | 2(5.9) | 0.436 |
| **Antibiotic** | 52(100) | 56(100) | 34(100) | |
| Azithromycin | 48(92.3) | 47(83.9) | 28(82.4) | 0.311 |
| Vancomycin | 48(92.3) | 53(94.6) | 29(85.3) | 0.294 |
| Piperacillin/Tazobactam | 45(86.5) | 47(83.9) | 21(61.8) | *0.012* *\*#* |
| Ceftriaxone | 37(71.2) | 40(71.4) | 19(55.9) | 0.246 |
| Cefepime | 14(26.9) | 22(39.3) | 11(32.4) | 0.392 |
| Amikacin | 14(26.9) | 23(41.1) | 18(52.9) | *0.048* *\** |
| **Antifungal** | 32(61.5) | 48(85.7) | 27(79.4) | 0.012 |
| Micafungin | 22(42.3) | 37(66.1) | 25(73.5) | *0.006* *\*$* |
| Fluconazole | 14(26.9) | 33(58.9) | 10(29.4) | *0.001* *$#* |
| **Respiratory Culture** | 51(98.1) | 56(100) | 33(97.1) | 0.478 |
| Positive Bacteria | 28(54.9) | 49(87.5) | 22(66.7) | *0.001* *$#* |
| Staphylococcus aureus | 11(21.6) | 12(21.4) | 5(15.2) | 0.728 |
| MRSA | 4(7.8) | 5(8.9) | 0(0) | 0.221 |
| Klebsiella pneumoniae | 2(3.9) | 8(14.3) | 2(6.1) | 0.135 |
| **Blood Culture** | 52(100) | 56(100) | 34(100) | |
| Positive Bacteria | 7(13.5) | 17(30.4) | 9(26.5) | 0.101 |
| **Hospitalization Data** | | | | |
| Hospital Length of Stay | 40[33–47] | 60[53–82] | 34[23–53] | *2.3968E-1.* *#$* |
| ICU Admission Day | 2[1–3] | 2[0–4] | 3[1–6] | 0.413 |
| Sampling Day | 10[6–14] | 13[8–16] | 13[8–16] | 0.115 |
| ICU Length of Stay | 28[21–33] | 52[41–63] | 29[21–40] | *7.3978E-12* *#$* |
| Intubation Day | 2[1–4] | 3[1–5] | 4[2–8] | 0.125 |
| Ventilator Days | 21 [16–24] | 41 [34–57] | 25[18–32] | *1.660 IE-19* *#$* |
| Average Follow-Up | 234[230–240] | 230[224–235] | - | *0.004* *$* |
| Days Between Death and ICU Admission | n.a. | n.a. | 30[22–47] | |

Data expressed as n(%) or median[interquartile range].

p-values denotes chi square and Kruskal-Wallis for categorical and continuous variables, respectively.

MV = Invasive mechanical ventilation

AAM = African American

BMI = The body mass index is the weight in kilograms divided by the square of the height in meters.

CHF = congestive heart failure, CAD = coronary artery disease, CKD = chronic kidney disease,

CVA = cerebrovascular accident

¥Biomarkers calculated as median value day 1–14 after initiation of mechanical ventilation

IL- 6 = interleukin 6, WBC = white blood cell count, CRP = c-reactive protein, PaO2 = partial pressure of arterial oxygen, FiO2 = fraction inspired oxygen

ECMO = extra corporeal membrane oxygenation

Anticoagulation = Full dose anticoagulation with therapeutic anti-Xa level>0.3IU/ml and/or PTT>45 sec.

Respiratory Culture = defined as having any respiratory culture performed

Positive Bacteria = a culture resulting in any bacteria growth

MRSA = methicillin-resistant Staphylococcus aureus

Blood Culture = defined as having any blood culture performed

ICU Admission Day = the number of days between hospital admission and ICU admission

Sampling Day = the number of days between hospital admission and day of sample collection

Intubation Day = number of days between hospital admission and day of intubation

Ventilator Days = total number of days on mechanical ventilation

Average Follow-Up = number of days between hospital admission and the last day of active follow-up

n.a. = not applicable.

Note.

* = significance between " 28 days" and "Deceased",

# = significance between ">28 days" and "Deceased",

$ = significance between " 28 days" and ">28 days"