medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

# Title: Impaired immune signaling and changes in the lung microbiome precede secondary bacterial pneumonia in COVID-19

**Authors:**

†Alexandra Tsitsiklis[1], †Beth Shoshana Zha[2], †Ashley Byrne[3], †Catherine DeVoe[1], †Elze Rackaityte[4], Sophia Levan[5], Sara Sunshine[5], Eran Mick[1,2,3], Rajani Ghale[1,2], Christina Love[1,2], Alexander J. Tarashansky[3], Angela Pisco[3], Jack Albright[3], Alejandra Jauregui[2], Aartik Sarma[2], Norma Neff[3], Paula Hayakawa Serpa[1], Thomas J. Deiss[2], Amy Kistler[3], Sidney Carrillo[2], K. Mark Ansel[6,7], Aleksandra Leligdowicz[2], Stephanie Christenson[2], Angela Detweiler[3], Norman G. Jones[8], Bing Wu[9], Spyros Darmanis[9], Susan V. Lynch[10,11], Joseph L. DeRisi[3,4], COMET Consortium+, Michael A. Matthay[2], Carolyn M. Hendrickson[2], Kirsten N. Kangelaris[5], Matthew F. Krummel[12], Prescott G. Woodruff[2,7], David J. Erle[5,13,14], Oren Rosenberg[1], Carolyn S. Calfee[2], *Charles R. Langelier[1,3]

†equal contributions

**Affiliations:**

[1] Department of Medicine, Division of Infectious Diseases, University of California San Francisco, San Francisco, CA, USA

[2] Department of Medicine, Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, University of California San Francisco, San Francisco, CA, USA

[3] Chan Zuckerberg Biohub, San Francisco, CA, USA

[4] Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, CA, USA

[5] Department of Medicine, University of California San Francisco, San Francisco, CA, USA

[6] Department of Microbiology and Immunology, University of California, San Francisco, CA, USA



NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[7] Sandler Asthma Basic Research Center, University of California, San Francisco, CA, USA

[8] Department of Experimental Medicine, University of California, San Francisco, CA, USA

[9] Genentech, Inc. San Francisco, CA, USA.

[10] Department of Gastroenterology, University of California, San Francisco, CA, USA

[11] Benioff Center for Microbiome Medicine, University of California, San Francisco, CA, USA

[12] Department of Pathology, University of California, San Francisco, CA, USA

[13] Lung Biology Center, University of California, San Francisco, CA, USA

[14] UCSF CoLabs, University of California, San Francisco, CA, USA

[+]COMET (COVID-19 Multiphenotyping for Effective Therapies) Consortium members are listed in the Supplementary Appendix.

*Corresponding author: Charles R. Langelier

Address: 499 Illinois Street, San Francisco, CA 94058

Phone: 801-201-5049

Email: chaz.langelier@ucsf.edu

Conflict of interest: The authors have declared that no conflict of interest exists.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Abstract**

Secondary bacterial infections, including ventilator-associated pneumonia (VAP), lead to worse clinical outcomes and increased mortality following viral respiratory infections including in patients with coronavirus disease 2019 (COVID-19). Using a combination of tracheal aspirate bulk and single-cell RNA sequencing we assessed lower respiratory tract immune responses and microbiome dynamics in 23 COVID-19 patients, 10 of whom developed VAP, and eight critically ill uninfected controls. At a median of three days (range: 2-4 days) before VAP onset we observed a transcriptional signature of bacterial infection. At a median of 15 days prior to VAP onset (range: 8-38 days), we observed a striking impairment in immune signaling in COVID-19 patients who developed VAP. Longitudinal metatranscriptomic analysis revealed disruption of lung microbiome community composition in patients with VAP, providing a connection between dysregulated immune signaling and outgrowth of opportunistic pathogens. These findings suggest that COVID-19 patients who develop VAP have impaired antibacterial immune defense detectable weeks before secondary infection onset.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## Introduction

Secondary bacterial pneumonia results in significant morbidity and mortality in patients with viral lower respiratory tract infections (LRTI)[1]. This problem was evident in the 1918 influenza pandemic during which the majority of deaths were ultimately attributed to secondary bacterial pneumonia[2]. SARS-CoV-2 infection, like influenza, confers an increased risk of late onset secondary bacterial infection, often manifesting as ventilator-associated pneumonia (VAP)[3,4]. Marked heterogeneity exists, however, with respect to the risk of VAP in patients with coronavirus disease 2019 (COVID-19), with incidence ranging from 12-87% in published cohort studies[5–8].

The mechanisms underlying VAP susceptibility in COVID-19 remain unknown. Animal models of influenza may provide some insight, suggesting a role for interferon-mediated suppression of cytokines essential for bacterial defense, including those important for neutrophil recruitment, antimicrobial peptide production and the Th17 response[9–11]. The impact of interferon signaling on antibacterial host defense, however, varies based on bacterial pathogen, and is not well studied in humans[12].

Much of our understanding of VAP immunobiology derives from sepsis studies, which describe a state of immune reprogramming in patients who go on to develop nosocomial infections[13–16]. This state is characterized by impaired monocyte, neutrophil and CD4 lymphocyte function, as well as decreased responsiveness of pathogen recognition receptors and downstream signaling pathways[13–16]. Severe COVID-19, like sepsis, is characterized by a dysregulated host response to infection, raising the possibility that similar mechanisms may contribute to the development of VAP in patients infected with SARS-CoV-2.

In addition to host immune elements, microbial factors contribute to VAP pathogenesis. Mechanically ventilated patients endure prolonged exposure to the hospital environment, which increases risk of colonization by opportunistic pathogens in the upper airway, oropharynx, and lungs[14,16]. Micro-aspiration or direct invasion may subsequently lead to lower respiratory tract disease and clinical onset of VAP[16,17]. Further, the common practice of early empiric antimicrobial

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

administration in critically ill patients can lead to disruption of the lung microbiome, which has been associated with the development of VAP[18,19].

Despite their interconnected roles, no studies to date have simultaneously profiled host immune responses and lung microbiome dynamics in the context of VAP. Further, the question of whether host responses to SARS-CoV-2 or other viral infections may contribute to this dysbiosis, leading to VAP, remains unanswered.

Given the marked heterogeneity in VAP incidence among patients with COVID-19[5–8], as well as gaps in mechanistic understanding of secondary bacterial pneumonia, we sought to assess the molecular determinants of VAP in patients infected with SARS-CoV-2. We employed a systems biology approach involving immunoprofiling the host transcriptional response and simultaneously assessing lung microbiome dynamics, using a combination of bulk and single cell RNA sequencing and extensive clinical phenotyping. We observed a striking impairment in antibacterial immune signaling at the time of intubation, that correlated with disruption of the lung microbiome, weeks before the onset of VAP.

## Results

We conducted a prospective case-control study of adults requiring mechanical ventilation for COVID-19 or for illnesses other than pneumonia. Of 63 patients with COVID-19 initially enrolled, tracheal aspirate (TA) specimens from 23 patients met inclusion criteria for primary analysis (**Methods, Figure 1)**. In addition, eight critically ill patients without COVID-19 from a second cohort (Study 2, **Methods**) were included as controls. Patients were enrolled at one tertiary care hospital and one safety net hospital in San Francisco, California under research protocols approved by the University of California San Francisco Institutional Review Board **(Methods)**. We collected TA periodically following intubation and performed bulk and single cell RNA-seq (scRNA-seq) **(Methods)**.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Patients with VAP were adjudicated using the United States Centers for Disease Control (CDC) definition[20], including a requirement for a positive bacterial TA culture (n=10). We defined onset of VAP as the first day a patient developed any of the criteria used to meet the definition, in accordance with CDC guidance[20]. Patients who did not meet the CDC criteria for VAP, and for whom there was no sustained clinical suspicion for bacterial pneumonia during the admission, were adjudicated as No-VAP (n=13). Patients who met CDC VAP criteria but had negative bacterial TA cultures, patients who were clinically suspected of having VAP with either positive or negative TA cultures, and patients who met clinical pneumonia criteria on intubation were excluded from this study **(Figure 1)**.

We compared lower respiratory tract host transcriptional responses between the VAP and No-VAP groups at two timepoints. “Early” timepoint TA samples were collected a median of two days post-intubation and 17 days before VAP onset (range: 15-38 days) for bulk RNA-seq analyses or 10 days before VAP onset (range: 8-16 days) for scRNA-seq analyses. “Late” timepoint samples were collected a median of two days before VAP onset (range: 2-3 days) for bulk RNA-seq analyses and three days before VAP onset (range: 2-4 days) for scRNA-seq analyses, and compared against samples collected from No-VAP patients at similar timepoints post-intubation (**Figure 1, Table 1**). No samples were analyzed after the onset of VAP. We additionally evaluated eight intubated patients with non-pneumonia illnesses as controls at the “early” timepoint.

We considered that sex as a biologic variable is important for acute COVID-19 severity, and thus asked whether differences in sex might explain our observations. However, no differences in sex were present between VAP and No-VAP groups (P=1, early and late timepoints) **(Table 1)**. There were also no significant differences between groups with respect to age, race or ethnicity **(Table 1, S1, S2)**. In addition, there were no differences between groups with respect to in-hospital receipt of any immunosuppressant or antibiotics prior to sample collection (**Table S1, S2, S3**).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**COVID-19 VAP is associated with a transcriptional signature of bacterial infection two days before VAP onset**

We began by assessing the lower respiratory host transcriptional response at the "late" timepoint, just before VAP onset. We carried out differential gene expression analysis on TA bulk RNA-seq data from five patients who developed VAP and eight No-VAP patients **(Table 1)**. We identified 436 differentially expressed genes at a False Discovery Rate (FDR) < 0.1 **(Figure 2A)** and performed gene set enrichment analysis (GSEA) **(Figure 2B)**.

As expected, the VAP group exhibited upregulation of pathways related to antibacterial immune responses in the days preceding secondary infection, such as neutrophil degranulation, toll-like receptor signaling, cytokine signaling and antigen presentation **(Figure 2B)**. We also noted that interferon alpha/beta signaling was upregulated. Ingenuity pathway analysis (IPA) predicted broad activation of upstream inflammatory cytokines (including IFN$\alpha$ and IFN$\gamma$) in the VAP group at this "late" timepoint **(Figure 2C)**.

**COVID-19 patients who develop VAP have attenuated immune signaling two weeks before VAP onset**

Given our findings of a unique lower respiratory transcriptional signature in the days preceding VAP onset, we next asked whether differences in host immune signaling might exist even earlier, two or more weeks before clinical diagnosis of VAP. We hypothesized that such differences might provide insight regarding the increased susceptibility to secondary bacterial infection in these patients.

We identified 154 differentially expressed genes at the "early" timepoint **(Table 1)**. The patients who developed VAP exhibited lower expression of several genes with established roles in innate immunity including *IFI30, MMP2, TLR9*, and *DEFB124* **(Figure 3A)**. GSEA further revealed that patients who developed VAP had lower expression of pathways related to antibacterial immune responses including neutrophil degranulation, toll-like receptor signaling, IL-

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

17 signaling, antigen presentation and complement pathways, and higher expression of interferon-alpha/beta signaling pathways, more than two weeks before the onset of VAP **(Figure 3B)**. Additionally, pathways related to adaptive immunity such as T and B cell receptor signaling were downregulated in the COVID-19 patients who later developed VAP **(Figure 3B)**.

To gauge the degree of immune pathway suppression compared to controls, we performed a similar analysis on critically ill intubated patients without infection **(Figure 3C)**. Relative to the control patients, multiple antibacterial immune pathways were downregulated in both COVID-19 groups, with the greatest attenuation in the VAP group **(Figure 3C)**. IPA upstream cytokine analysis predicted reduced activation of multiple proinflammatory cytokines juxtaposed against increased activation of IFNB1 in the VAP patients **(Figure 3D)**. Compared to the control group, the predicted activation states of several proinflammatory cytokines were decreased in both COVID-19 groups **(Figure S2)**.

Given prior reports demonstrating a correlation between SARS-CoV-2 viral load and interferon-related gene expression[21] we next asked whether viral load differed between VAP and No-VAP patients. At the "early" timepoint, no differences were observed in either PCR cycle threshold or bulk RNA-seq SARS-CoV-2 reads per million (P=0.53 and P=0.84, respectively) **(Figure S3)**. We also considered the possibility that differences in the number of days of steroid exposure prior to sample collection might explain our findings but found no differences between groups (P=0.34) **(Table S1, S3)**.

**COVID-19 VAP is associated with impaired antibacterial immune signaling in monocytes and macrophages**

To further understand the observed downregulation of key antibacterial signaling pathways in COVID-19 patients who developed VAP, we next asked whether this was driven by specific types of immune cells. We performed scRNA-seq on TA specimens and enriched for immune cells using CD45 selection **(Methods)**. We analyzed 12,197 cells from 14 patients after

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

quality control and batch correction. First, we analyzed samples obtained post-intubation at the "early" timepoint. Clustering based upon cellular transcriptional signatures followed by a comparison of cell type proportions did not reveal statistically significant differences between VAP and No-VAP groups with respect to any cell type at the "early" timepoint **(Figure 4A, S4**). Monocytes and macrophages were the most abundant cell types, and thus we focused transcriptional assessment on these two populations.

COVID-19 patients who developed VAP had distinct post-intubation transcriptional signatures compared to the No-VAP patients at this "early" timepoint, with respect to the overall monocyte/macrophage cluster **(Figure 4, S4, S5)**. We identified 625 differentially expressed genes (FDR < 0.05) in this cluster and noted downregulation of several genes with key roles in innate immunity in the COVID-19 patients who subsequently developed VAP compared to those who did not, including *IL1Rn, CXCL2, CCL20*, and *JunB*. (**Figure 4C**). In addition, we noted upregulation of interferon-induced genes including *IFI30* and *IFNγR-2*. **(Figure 4C)**. Findings were similar when evaluating differential expression within alveolar macrophages, the largest monocyte/macrophage subset (**Figure S5**).

IPA canonical pathway analysis revealed downregulation of several cytokine and innate immune signaling pathways in the patients who later developed VAP at the "early" post-intubation timepoint, including IL-6 and IL-1 signaling, and the NRF2-mediated oxidative stress response (**Figure 4D)**. Computational analysis of upstream cytokine activation predicted impaired activation of multiple pro-inflammatory cytokines in patients who developed VAP, including TNF, IFNγ, and IL1B **(Figure 4E)**. Similar findings were observed in the alveolar macrophage population **(Figure S5)**. We were unable to reliably assess the T cell population due to low cell numbers.

A comparison of monocytes and macrophages between the VAP and No-VAP patients just prior to VAP onset at the "late" timepoint revealed significant upregulation of genes essential for antibacterial defense. These findings reflected bulk RNA-seq observations and were

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

characterized by upregulation in the VAP group of pathways related to IL-17 and nitric oxide signaling, reactive oxygen species production, and hypercytokinemia **(Figure S6)**.

**Temporal dynamics of the host response in COVID-19 patients who develop VAP**

We next investigated temporal dynamics of the lower airway host immune response in COVID-19 patients from the time of intubation to development of VAP, by evaluating bulk RNA-seq differential gene expression between COVID-19 VAP patients at the "early" timepoint (median of 17 days before VAP onset, n=4) versus "late" timepoint (median of two days before VAP onset, n=5). We identified 2705 differentially expressed genes, and unsupervised hierarchical clustering of the 50 most significant genes demonstrated clear separation of the two timepoints **(Figure 5A)**. GSEA revealed that type I interferon signaling was notably downregulated at the "late" timepoint in comparison to the "early" timepoint **(Figure 5B)**; however, expression of this pathway was still significantly higher in the VAP versus the No-VAP patients **(Figure 2B)**. Several other immune signaling pathways were more highly expressed at the "late" timepoint, presumably reflecting activation of an antibacterial response in the setting of impending bacterial pneumonia **(Figure 5B)**. Consistent with this, upstream regulator analysis predicted increased activation of several pro-inflammatory cytokines and decreased IFN$\alpha$ and IFN$\lambda$ signaling at the "late" versus "early" timepoints **(Figure 5C)**.

In contrast, comparing No-VAP patients at the "early" (n=8) versus "late" (n=8) timepoints yielded only two genes, both of which were interferon-stimulated genes (*RSAD2* and *CMPK2*) downregulated at the "late" timepoint, suggesting that while the host response was relatively unchanged in these patients, the antiviral response attenuated over time. Indeed, GSEA revealed that type I interferon signaling, and other antiviral immune pathways, were downregulated in the patients who did not develop VAP at the "late" timepoint **(Figure S7)**.

We further compared dynamics of host immune responses between VAP and No-VAP patients by performing longitudinal analyses of key immune signaling pathways, including all

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

patients with available bulk RNA-seq samples (VAP: n=7, No-VAP: n=10). Onset of VAP ranged from 10-39 days post intubation (median 19 days), and treatment with immunosuppressants did not differ significantly between comparator groups (p=0.30, Fisher's exact test). Early attenuation of immune signaling in the VAP group was conspicuous, and this pattern eventually resolved later in disease course by the time secondary bacterial infection became established **(Figures 5D-G)**. We confirmed that the observed differences between VAP and No-VAP patients were not driven by differences in treatment with immunosuppressants by comparing pathway Z-scores between patients who received immunosuppressants and those who did not, at the "early" timepoint **(Figure S8)**.

We subsequently performed a similar comparison between the "early" and "late" timepoints based on scRNA-seq data from patients who developed VAP. Differential gene expression analysis between these two populations identified 1466 differentially expressed genes in the encompassing monocyte/macrophage cluster. IPA on these genes revealed upregulation of antibacterial signaling pathways at the "late" timepoint, including signaling by several cytokines (including IL-17, IL-6, TNF, IFN) **(Figure 5H-I)**, congruent with the bulk RNA-seq analysis.

**Lung microbiome disruption precedes VAP in COVID-19 patients**

We hypothesized that the innate immune suppression observed in patients who developed VAP would correlate with SARS-CoV-2 viral load. Using TA metatranscriptomics to assess the lower respiratory microbiome, we evaluated longitudinal changes in SARS-CoV-2 abundance. Although no difference in viral load was observed at the "early" timepoint **(Figure S3, Figure 6A)**, the trajectory of SARS-CoV-2 viral load differed significantly in patients who developed VAP. Viral load decreased over time in both groups, however the virus was still detected at later timepoints in patients with VAP, while it was undetected in any of the No-VAP patients at these timepoints (**Figure 6A**). This result suggested that COVID-19 patients who develop VAP may exhibit

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

impaired ability to clear the virus compared to those who do not, and that the lung microbiome composition may be similarly impacted.

Indeed, COVID-19 patients who developed VAP exhibited a significant reduction in bacterial alpha diversity of their airway microbiome before clinical signs of infection (Shannon Diversity Index, P=0.0065; **Figure 6B**). Beta diversity did not differ significantly, which may be due to the fact that VAP was caused by different pathogens in each patient, resulting in different airway microbiome compositions in each case **(Figure S9)**. All patients received antibiotics prior to collection of the first sample, suggesting that antibiotic use was not driving these differences **(Table S1)**.

We next assessed the most abundant bacterial genera in the lung, as identified by TA metatranscriptomics. At the "early" timepoint, control and No-VAP samples were predominated by common lung commensals, with significant enrichment (FDR < 0.05) of *Alloprevotella*, *Fusobacterium*, and *Neisseria* compared to the VAP group (**Figure 6C**). In the VAP group, early significant enrichment of *Klebsiella* was noted, suggestive of a shift in the microbiome structure that precedes disease onset and culture positive confirmation (**Figure 6C**). By the "late" timepoint, VAP-associated pathogens such as *Citrobacter*, *Escherichia*, and *Klebsiella* were highly prevalent. *Citrobacter* was significantly enriched in VAP versus No-VAP groups at this timepoint, supporting culture results. While highly enriched in VAP, *Klebsiella* did not reach significance at this timepoint due to high variance between individuals (**Figure 6D**). No-VAP patients had a significantly greater abundance of commensal lung microbes, such as *Lactobacillus*, *Bifidobacterium*, and *Veillonella,* at the "late" timepoint **(Figure 6D)**.

## Discussion

Secondary bacterial pneumonia contributes to significant morbidity and mortality in patients with primary viral lower respiratory tract infections[1,3], but mechanisms governing individual susceptibility to VAP have remained unclear. Few human cohort studies have evaluated

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the immunologic underpinnings of VAP, and none have been reported in the context of COVID-19, which is characterized by a dysregulated host response distinct from other viral pneumonias[21–23]. To address this gap and probe mechanisms of VAP susceptibility in patients with COVID-19, we carried out a systems biological assessment of host and microbial dynamics of the lower respiratory tract.

Two days before VAP onset, a transcriptional signature consistent with bacterial infection was observed. This finding suggests that host response changes can occur before clinical recognition of pneumonia, highlighting the potential utility of the host transcriptome as a tool for VAP surveillance. While intriguing, this observation did not provide an explanation for differential susceptibility of some COVID-19 patients to post-viral pneumonia.

The discovery of an early suppressed antibacterial immune response in patients who later developed VAP did, however, offer a potential explanation. More than two weeks before VAP onset, we observed a striking suppression of pathways related to both innate and adaptive immunity, including neutrophil degranulation, toll-like receptor (TLR) signaling, complement activation, antigen presentation, and T cell receptor and B receptor signaling, as well as signaling by several cytokines (e.g. IL-1, IL-4, IL-12, IL-13 and IL-17). Comparison against uninfected, intubated controls confirmed the previously described paradoxical impairment in immune signaling found in patients with severe COVID-19[22], and suggested that VAP susceptibility may be the result of disproportionate suppression of innate and adaptive pathways critical for antibacterial defense, resulting in enhanced susceptibility to opportunistic secondary infections.

Animal models of influenza have provided insight into potential mechanisms of post-viral pneumonia, although none have provided insight regarding why some individuals are more susceptible than others. In mice inoculated with influenza, for instance, virus-induced type I interferon suppresses neutrophil chemokines and impairs Th17 immunity, compromising effective clearance of bacterial infections[10,11]. Interestingly, we also observed increased type I interferon signaling in COVID-19 patients who weeks later developed VAP, and an impairment in IL-17

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

signaling and other immune pathways. Desensitization to TLR ligands after influenza infection has also been documented[24], which is congruent with the downregulation of TLR signaling at the time of intubation observed in our bulk RNA-seq analyses.

We asked whether certain cell types were responsible for driving the early suppression of immune signaling observed in COVID-19 patients who went on to develop VAP. No significant differences in proportions of any cell types was observed between patients with or without VAP at the time of intubation. This finding suggests that an impairment of immune cell recruitment was not causing these differences, but rather significant gene expression differences within these immune cell populations.

In the most predominant cell population, the monocytes and macrophages, we observed broad downregulation of the innate and adaptive immune responses, concordant with global observations in bulk RNA-seq analyses. Further, we noted a downregulation of key pathways involved in antimicrobial immune responses including TREM1, IL-1, and IL-6 signaling. Overall, our data suggest that while no difference in cell type populations existed between groups, changes in the gene expression within monocytes and macrophages contributes to immune suppression in COVID-19 patients who later develop VAP.

SARS-CoV-2 viral load correlates with interferon stimulated gene expression[21,22] and thus we initially hypothesized that differences in viral load between groups might relate to individual VAP susceptibility. However, we found no difference between groups at the “early” timepoint. Moreover, no differences existed in terms of immunosuppressive medication administration or clinically diagnosed immunodeficiency, suggesting that other, still unidentified mechanisms present at the time of intubation must underlie the marked suppression of immune gene expression in COVID-19 patients who went on to develop VAP.

While no difference in viral load was observed at the time of intubation, the COVID-19 patients who developed VAP exhibited impaired viral clearance over the time-course of intubation. This observation was corroborated by a prolonged antiviral type I interferon response at the “late”

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

timepoint (median of two days before VAP onset) in patients who developed VAP versus those who did not, pointing to the persistence of suboptimal antiviral immunity in these patients. Early induction of functional SARS-CoV-2 specific T cells is associated with faster viral clearance in COVID-19 patients[25] and likewise, we observed impairments in T cell activation and signaling in the VAP group by bulk RNA-seq at the "early" timepoint, which further suggests a decreased ability to control the virus in these patients.

Respiratory viruses can reshape the human airway microbiome by modulating host inflammatory responses[26,27]. In mouse models of influenza, the airway microbiome exhibits expansion of several bacterial families during the course of viral infection as innate immunity is suppressed[26]. These changes increase the risk of secondary bacterial infection[26] and have been observed in patients with chronic obstructive pulmonary disease, where suppression of the innate immune response in rhinovirus infected patients may be followed by bacterial superinfection[28,29].

Similarly, the innate immune suppression observed in COVID-19 patients who developed VAP was associated with airway microbiome collapse and the outgrowth of lung pathogens in advance of clinical VAP diagnosis. This finding suggests that individual immune responses to SARS-CoV-2 infection may drive a restructuring of the microbial community and increase susceptibility to VAP (**Figure 7**). The resulting outgrowth of a VAP-associated bacterial pathogen may elicit an antibacterial response, but the broader immunosuppressive state preceding this response may be insufficient to control the development of clinical pneumonia. Those with a lesser degree of immunosuppression may be able to respond faster and therefore control opportunistic bacterial pathogens more effectively.

These findings may also have important implications for management of patients with COVID-19 related acute respiratory failure, many of whom are now being treated with corticosteroids and other immunosuppressive drugs. These agents may lead to further suppression of the key pathways required for host response to secondary bacterial infection. Thus, our results emphasize the need for ongoing vigilance for VAP in patients treated with potent

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

immunosuppressive agents, as well as the need to develop novel diagnostic and/or prognostic approaches to identifying patients at highest risk.

Protein biomarkers of immune function have been associated with nosocomial infection in critically ill patients[15], although not in the context of viral infection. Further studies with a larger cohort may allow for the identification of transcriptomic biomarkers to assess a COVID-19 patient's risk of VAP at the time of intubation, which could reduce inappropriate use of prophylactic antibiotics or immunomodulatory treatments, or signal a need for enhanced surveillance strategies.

Sample size is a limitation of this study, which could result in over-fitting to this small set of patients; however, the reproducibility of our observations across both bulk and scRNA-seq analyses and the significant number of differentially expressed genes among the comparator groups support the internal validity of our conclusions. Additional studies in a larger independent cohort will be needed to determine the utility of lower respiratory gene expression biomarkers for predicting future development of VAP. Because this study was limited to critically ill, intubated patients, we were unable to assess early stages of COVID-19, which may provide additional insight regarding determinants of secondary bacterial infection.

We were unable to assess whether epithelial cells contributed to VAP risk due to enrichment for immune cells prior to scRNA-seq. However, given that monocytes, macrophages and lymphocytes can both be infected by SARS-CoV-2 and contribute to disease pathophysiology and resolution[30–32], our findings are apropos. With larger cohorts, the early detection of specific immune pathway suppression and microbiome collapse could be leveraged to develop clinically useful models for identifying COVID-19 patients with increased susceptibility to secondary bacterial pneumonia.

## Materials and Methods

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Study design, cohorts, and enrollment**

We conducted a prospective case-control study of adults requiring mechanical ventilation for COVID-19 with or without secondary bacterial pneumonia. We also evaluated control patients requiring mechanical ventilation for other reasons who had no evidence of pulmonary infection (**Figure 1**). Patients were enrolled in either of two prospective cohort studies of critically ill patients at the University of California, San Francisco (UCSF) and Zuckerberg San Francisco General Hospital between 07/2013 and 07/2020[33]. Of the COVID-19 patients, 17 were co-enrolled in the National Institute of Allergy and Infectious Diseases-funded Immunophenotyping Assessment in a COVID-19 Cohort (IMPACC) Network study.

**Ventilator-associated pneumonia adjudication**

A total of 63 adults who required intubation for severe COVID-19 (Cohort 1) and who had available TA samples were considered for inclusion in the study (**Figure 1**). Patients who met the US Centers for Disease Control (CDC) definition for VAP[20] (based on the PNU1 and PNU2 definitions) and had a positive bacterial respiratory culture were adjudicated as having VAP for the purpose of the study (n=16). Patients who did not meet these criteria, and for whom there was no sustained clinical suspicion for bacterial pneumonia during the admission, were categorized as No-VAP (n=17).

VAP and No-VAP patients for whom samples at the timepoints of interest were available were included in the analyses (VAP: n=10; No-VAP: n=13). Patients who met CDC-VAP criteria but had negative TA cultures (n=5 patients), as well as patients with clinically suspected bacterial pneumonia who did not meet CDC VAP criteria and had a positive (n=6) or negative (n=5) TA culture were excluded. Fourteen patients who met clinical pneumonia criteria on intubation were also excluded. Eight intubated patients from a recent study[22] (Cohort 2) were included as controls and were selected because they had previously been adjudicated as having no evidence of lower respiratory tract infection. This group included four patients with acute respiratory distress

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

syndrome (ARDS) due to non-infectious etiologies, and four patients without ARDS who were intubated for other reasons (subdural hematoma (n=1), retroperitoneal hemorrhage (n=1), or neurosurgical procedures (n=2)).

**Tracheal aspirate sampling**

Following enrollment, tracheal aspirate (TA) was collected (periodically following intubation for Study 1, or once within three days of intubation for Study 2), without addition of saline wash, and either a) mixed 1:1 with DNA/RNA shield (Zymo Research) for bulk RNA-seq or b) immediately processed in a biosafety level 3 laboratory (BSL3) for scRNA-seq analysis.

**Bulk RNA sequencing and host transcriptome analysis**

RNA sequencing

To evaluate host and microbial gene expression, metatranscriptomic next generation RNA sequencing (RNA-seq) was performed on TA specimens. Following RNA extraction (Zymo Pathogen Magbead Kit) and DNase treatment, human cytosolic and mitochondrial ribosomal RNA was depleted using FastSelect (Qiagen). To control for background contamination, we included negative controls (water and HeLa cell RNA) as well as positive controls (spike-in RNA standards from the External RNA Controls Consortium (ERCC))[34]. RNA was then fragmented and underwent library preparation using the NEBNext Ultra II RNA-seq Kit (New England Biolabs). Libraries underwent 146 nucleotide paired-end Illumina sequencing on an Illumina Novaseq 6000.

Host differential expression

Following demultiplexing, sequencing reads were pseudo-aligned with kallisto[35] to an index consisting of all transcripts associated with human protein coding genes (ENSEMBL v. 99), cytosolic and mitochondrial ribosomal RNA sequences and the sequences of ERCC RNA standards. Gene-level counts were generated from the transcript-level abundance estimates

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

using the R package tximport[36], with the scaledTPM method. Samples retained in the dataset had a total of at least 1,000,000 estimated counts associated with transcripts of protein coding genes.

Genes were retained for differential expression analysis if they had counts in at least 30% of samples. Differential expression analysis was performed using the R package DESeq2[37]. We modeled the expression of individual genes using the design formula ~VAPgroup, where VAP groups were “VAP-early”, “No-VAP-early”, “VAP-late” and “No-VAP-late” and used the results() function to extract a specific contrast. Separate comparisons to the control group were performed using the design formula ~COVID-19-status to compare positive and negative patients.

Significant genes were identified using a Benjamini-Hochberg false discovery rate (FDR) < 0.1. We generated heatmaps of the top 50 differentially expressed genes by FDR. For visualization, gene expression was normalized using the regularized log transformation, centered, and scaled prior to clustering. Heatmaps were generated using the *pheatmap* package. Columns were clustered using Euclidean distance and rows were clustered using Pearson correlation. Differential expression analysis results are provided in (**Supplementary Data File 1**).

Pathway analysis

Gene set enrichment analyses (GSEA) were performed using the fgseaMultilevel function in the R package fgsea[38] and REACTOME[39] pathways with a minimum size of 10 genes and a maximum size of 1,500 genes. All genes were included in the comparison, pre-ranked by the test statistic. Significant pathways were defined as those with a Benjamini-Hochberg adjusted P-value < 0.05. Ingenuity Pathway Analysis (IPA) Canonical Pathway and Upstream Regulator Analysis[40] was employed on genes with P < 0.1 and ranked by the test statistic to identify cytokine regulators. Significant IPA results were defined as those with a Z-score absolute value greater than 2 and an overlap P-value < 0.05. The gene sets in figures were selected to reduce redundancy and highlight diverse biological functions. Full GSEA and IPA results are provided in (**Supplementary Data Files 2 and 3**).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Longitudinal pathway analysis was performed using all available TA samples spanning post-intubation to VAP onset for all patients included in the bulk RNA-seq analysis. Analysis was restricted to samples with at least 1,000,000 human protein coding transcripts. Pathways of interest were selected from the significant GSEA results of the comparison of VAP vs. No-VAP patients in the "early" timepoint. The top 20 leading edge genes were selected from each pathway for analysis. To calculate a Z-score for each gene, expression was normalized using the variance stabilizing transformation (VST), centered, and scaled. A pathway Z-score was calculated by averaging the 20 gene Z-scores. Multiple Z-scores per patient at a given time interval were averaged so that each patient corresponds to one datapoint at each interval. Statistical significance of pathway expression over time between VAP and No-VAP groups was calculated using a two-way analysis of variance (ANOVA) in GraphPad PRISM.

**Single cell RNA sequencing and transcriptome analysis**

After collection, fresh TA was transported to a BSL-3 laboratory at ambient temperature to improve neutrophil survival. 3mL of TA was dissociated in 40mL of PBS with 50ug/mL collagenase type 4 (Worthington) and 0.56 ku/mL of Dnase I (Worthington) for 10 minutes at room temperature, followed by passage through a 70μM filter. Cells were pelleted at 350g 4C for 10 minutes, resuspended in PBS with 2mM EDTA and 0.5% BSA, and manually counted on a hemocytometer. Cells were stained with MojoSort Human CD45 and purified by the manufacturer's protocol (Biolegend). After CD45 positive selection, cells were manually counted with trypan blue on a hemocytometer. Using a V(D)J v1.1 kit according to the manufacturer's protocol, samples were loaded on a 10X Genomics Chip A without multiplexing, aiming to capture 10,000 cells (10X Genomics). Libraries underwent paired end 150 base pair sequencing on an Illumina NovaSeq6000.

Raw sequencing reads were aligned to GRCh38 using the STAR aligner . Cell barcodes were then determined based upon UMI count distribution. Read count matrices were generated

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

through the 10X genomics cellranger pipeline v3.0. Data was processed and analyzed using the Scanpy v1.6[42]. Cells that had <200 genes, <20% mitochondrial reads, and had greater than 2500 counts were filtered. Mitochondrial genes were removed and multi-sample integration was performed using single-cell variational inference (scVI)[43]. Differential expression was performed using MAST v1.16.0[44]. Differential expression analysis results are detailed in (**Supplementary Data File 4**).

Pathway analysis

Ingenuity Pathway Analysis (IPA) Canonical Pathway and Upstream Regulator Analysis[40] was employed on genes with $P < 0.05$ and ranked by log2foldchange to identify canonical pathways and cytokine regulators. Due to the significantly greater number of differentially expressed genes in scRNA-seq analyses, we utilized a more restrictive cutoff of $FDR < 0.05$ for scRNA-seq to ensure a similar number of genes were input into IPA. Significant IPA results were defined as those with a Z-score absolute value greater than 2 and an overlap P-value < 0.05. The gene sets in figures were selected to reduce redundancy and highlight diverse biological functions. Full GSEA and IPA results are provided in (**Supplementary Data Files 5 and 6**).

**Lung microbiome analysis**

RNA from tracheal aspirates was sequenced as described above. Taxonomic alignments were obtained from raw sequencing reads using the IDseq pipeline[45,46], which performs quality filtration and removal of human reads followed by reference-based taxonomic alignment at both the nucleotide and amino acid level against sequences in the National Center for Biotechnology Information (NCBI) nucleotide (NT) and non-redundant (NR) databases, followed by assembly of reads matching each taxon detected. Taxonomic alignments underwent background correction for environmental contaminants (see below), viruses were excluded, and data was then aggregated to the genus level before calculating diversity metrics. Alpha diversity (Shannon's

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Diversity Index) and beta diversity (Bray-Curtis dissimilarity) were calculated and the latter plotted using non-metric multidimensional scaling (NDMS). Comparison of alpha and beta diversity over time between VAP and No-VAP groups was calculated using a two-way analysis of variance (ANOVA) in R. Top ten taxa in each group were identified by ranking the percent of counts for each genus among total bacterial counts in each group. Significantly enriched taxa were evaluated by comparing distributions using three models (Poisson, negative binomial, and zero-inflated negative binomial mixed-effects models) and correcting for multiple hypothesis testing using Benjamini-Hochberg as previously described[47]. Significantly enriched taxa were called using a $q < 0.05$ cutoff.

Identification and mitigation of environmental contaminants

To minimize inaccurate taxonomic assignments due to environmental and reagent derived contaminants, non-templated “water only” and HeLa cell RNA controls were processed with each group of samples that underwent nucleic acid extraction. These were included, as well as positive control clinical samples, with each sequencing run. Negative control samples enabled estimation of the number of background reads expected for each taxon. A previously developed negative binomial model[21] was employed to identify taxa with NT sequencing alignments present at an abundance significantly greater compared to negative water controls. This was done by modeling the number of background reads as a negative binomial distribution, with mean and dispersion fitted on the negative controls. For each batch (sequencing run) and taxon, we estimated the mean parameter of the negative binomial by averaging the read counts across all negative controls, slightly regularizing this estimate by including the global average (across all batches) as an additional sample. We estimated a single dispersion parameter across all taxa and batches, using the functions glm.nb() and theta.md() from the R package MASS[48]. Taxa that achieved a P-value < 0.01 were carried forward.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Statistics**

Statistical significance was defined as $P < 0.05$, using two-tailed tests. Categorical data were analyzed by Fisher's exact test and nonparametric continuous variables were analyzed by Mann-Whitney. Statistical approaches used for gene expression and microbiome analyses are detailed in each respective Methods section.

**Study approval**

Both cohort studies were approved by the UCSF Institutional Review Board (IRB) under protocols 10-02701 (control patients, pre-COVID-19 pandemic) and 20-30497 (COVID-19 patients, COVID-19 Multiphenotyping for Effective Therapy (COMET) study), respectively. Detailed enrollment and consent protocols for both studies have been previously described[18,33].

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

## References

1. Falsey, A. R. *et al.* Bacterial Complications of Respiratory Tract Viral Illness: A Comprehensive Evaluation. *The Journal of Infectious Diseases* **208**, 432–441 (2013).
2. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. *J Infect Dis* **198**, 962–970 (2008).
3. Rouzé, A. *et al.* Relationship between SARS-CoV-2 infection and the incidence of ventilator-associated lower respiratory tract infections: a European multicenter cohort study. *Intensive Care Med* 1–11 (2021) doi:10.1007/s00134-020-06323-9.
4. DeVoe, C. *et al.* Increased Rates of Secondary Bacterial Infections, Including Enterococcus Bacteremia, in Patients Hospitalized with COVID-19. *Infect Control Hosp Epidemiol* 1–33 (2021) doi:10.1017/ice.2021.391.
5. Luyt, C.-E. *et al.* Ventilator-associated pneumonia in patients with SARS-CoV-2-associated acute respiratory distress syndrome requiring ECMO: a retrospective cohort study. *Ann Intensive Care* **10**, 158 (2020).
6. Bardi, T. *et al.* Nosocomial infections associated to COVID-19 in the intensive care unit: clinical characteristics and outcome. *Eur J Clin Microbiol Infect Dis* (2021) doi:10.1007/s10096-020-04142-w.
7. Maes, M. *et al.* Ventilator-associated pneumonia in critically ill patients with COVID-19. *Crit Care* **25**, 25 (2021).
8. Søgaard, K. K. *et al.* Community-acquired and hospital-acquired respiratory tract infection and bloodstream infection in patients hospitalized with COVID-19 pneumonia. *J Intensive Care* **9**, (2021).
9. Metzger, D. W. & Sun, K. Immune Dysfunction and Bacterial Co-Infections following Influenza. *J Immunol* **191**, 2047–2052 (2013).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

10. Shahangian, A. *et al.* Type I IFNs mediate development of postinfluenza bacterial pneumonia in mice. *J Clin Invest* **119**, 1910–1920 (2009).
11. Kudva, A. *et al.* Influenza A inhibits Th17-mediated host defense against bacterial pneumonia in mice. *J Immunol* **186**, 1666–1674 (2011).
12. Boxx, G. M. & Cheng, G. The Roles of Type I Interferon in Bacterial Infection. *Cell Host Microbe* **19**, 760–769 (2016).
13. Pelekanou, A. *et al.* Decrease of CD4-lymphocytes and apoptosis of CD14-monocytes are characteristic alterations in sepsis caused by ventilator-associated pneumonia: results from an observational study. *Crit Care* **13**, R172 (2009).
14. Wunderink, R. G. Nosocomial Pneumonia, Including Ventilator-associatedPneumonia. *Proc Am Thorac Soc* **2**, 440–444 (2005).
15. Conway Morris, A. *et al.* Cell-surface signatures of immune dysfunction risk-stratify critically ill patients: INFECT study. *Intensive Care Med* **44**, 627–635 (2018).
16. Ballinger, M. N. & Standiford, T. J. Innate Immune Responses in Ventilator-Associated Pneumonia. in *Mucosal Immunology of Acute Bacterial Pneumonia* (ed. Prince, A.) 185–212 (Springer New York, 2013). doi:10.1007/978-1-4614-5326-0_8.
17. Kalil, A. C. *et al.* Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis* **63**, e61–e111 (2016).
18. Langelier, C. *et al.* Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. *Proc Natl Acad Sci U S A* **115**, E12353–E12362 (2018).
19. Flanagan, J. L. *et al.* Loss of bacterial diversity during antibiotic treatment of intubated patients colonized with Pseudomonas aeruginosa. *J Clin Microbiol* **45**, 1954–1962 (2007).
20. CDC-NHSN. Pneumonia (Ventilator-associated [VAP] and non-ventilator-associated Pneumonia [PNEU]) Event.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

21. Mick, E. *et al.* Upper airway gene expression reveals suppressed immune responses to SARS-CoV-2 compared with other respiratory viruses. *Nature Communications* **11**, 5854 (2020).
22. Sarma, A. *et al.* COVID-19 ARDS is characterized by a dysregulated host response that differs from cytokine storm and is modified by dexamethasone. *Res Sq* (2021) doi:10.21203/rs.3.rs-141578/v1.
23. Blanco-Melo, Daniel. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell.* doi:10.1016/j.cell.2020.04.026.
24. Didierlaurent, A. *et al.* Sustained desensitization to bacterial Toll-like receptor ligands after resolution of respiratory influenza infection. *J Exp Med* **205**, 323–329 (2008).
25. Tan, A. T. *et al.* Early induction of functional SARS-CoV-2-specific T cells associates with rapid viral clearance and mild disease in COVID-19 patients. *Cell Reports* **34**, 108728 (2021).
26. Goulding, J. *et al.* Lowering the Threshold of Lung Innate Immune Cell Activation Alters Susceptibility to Secondary Bacterial Superinfection. *J Infect Dis* **204**, 1086–1094 (2011).
27. Man, W. H., de Steenhuijsen Piters, W. A. A. & Bogaert, D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* **15**, 259–270 (2017).
28. Mallia, P. *et al.* Rhinovirus Infection Induces Degradation of Antimicrobial Peptides and Secondary Bacterial Infection in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **186**, 1117–1124 (2012).
29. Molyneaux, P. L. *et al.* Outgrowth of the Bacterial Airway Microbiome after Rhinovirus Exacerbation of Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **188**, 1224–1231 (2013).
30. The NU SCRIPT Study Investigators *et al.* Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. *Nature* **590**, 635–641 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

31. Borsa, M. & Mazet, J. M. Attacking the defence: SARS-CoV-2 can infect immune cells. *Nat Rev Immunol* **20**, 592–592 (2020).
32. Boumaza, A. *et al.* Monocytes and Macrophages, Targets of Severe Acute Respiratory Syndrome Coronavirus 2: The Clue for Coronavirus Disease 2019 Immunoparalysis. *J Infect Dis* **224**, 395–406 (2021).
33. Sarma, A. *et al.* Tracheal aspirate RNA sequencing identifies distinct immunological features of COVID-19 ARDS. *Nat Commun* **12**, 5152 (2021).
34. Pine, P. S. *et al.* Evaluation of the External RNA Controls Consortium (ERCC) reference material using a modified Latin square design. *BMC Biotechnol* **16**, 54 (2016).
35. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525–527 (2016).
36. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
37. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
38. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv* 060012 (2019) doi:10.1101/060012.
39. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res* **42**, D472-477 (2014).
40. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523–530 (2014).
41. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
42. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY : large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).
43. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep Generative Modeling for Single-cell Transcriptomics. *Nat Methods* **15**, 1053–1058 (2018).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

44. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, 278 (2015).

45. Ramesh, A. *et al.* Metagenomic next-generation sequencing of samples from pediatric febrile illness in Tororo, Uganda. *PLoS One* **14**, e0218318 (2019).

46. Kalantar, K. L. *et al.* IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* **9**, (2020).

47. Durack, J. *et al.* Delayed gut microbiota development in high-risk for asthma infants is temporarily modifiable by Lactobacillus supplementation. *Nat Commun* **9**, 707 (2018).

48. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer-Verlag, 2002). doi:10.1007/978-0-387-21706-2.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Author contributions:

Conceptualization: CRL, AT, BSZ, AB, CD, SL, ER, AP, OR, CSC

Methodology: CRL, AT, BSZ, AB, CD, SL, ER, AJT, OR

Data acquisition: RG, AJ, PHS, TJD, BSZ, AB, NGJ, AD

Formal analysis: AT, BSZ, AB, CD, SL, ER, AJT, EM, JA

Investigation: BSZ, AB, CD, SS, CSC, DJE

Funding acquisition: CRL, CSC, JLD, DJE

Supervision: CRL, OR, NN, JLD, CSC, DJE

Writing - original draft: AT, BSZ, AB, CD, SL, ER, CRL

Writing - review & editing: All authors

Acknowledgements: This study was performed with support from the National Institute of Allergy and Infectious Diseases-sponsored Immunophenotyping Assessment in a COVID-19 Cohort (IMPACC) Network. We gratefully appreciate support from Amy Kistler, PhD, Jack Kamm, PhD, Saharai Caldera, BS and Maira Phelps BS.

Funding: Funding for the COMET cohort enrollment, sample collection and data analysis derived from: K23HL138461-01A1 (CL), K24HL137013 (PGW), F32 HL151117 (AS), R35 HL140026 (CSC), NIAID U19AI077439 (DJE), Chan Zuckerberg Biohub (AB, JLD). The UCSF IMPACC site was funded by NIAID U19AI077439 (DJE). Funding for enrollment of COMET participants not enrolled in IMPACC and for all sample collection and data analysis derived from K23HL138461-01A1 (CL), K24HL137013 (PGW), F32 HL151117 (AS), R35 HL140026 (CSC), and the Chan Zuckerberg Biohub (AB, JLD). Philanthropic support was provided from Mark and Carrie Casey, Julia and Kevin Hartz, Carl Kawaja and Wendy Holcombe, Eric Keisman and Linda Nevin, Martin and Leesa Romo, Three Sisters Foundation, Diana Wagner and Jerry Yang and Akiko Yamazaki.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Data and materials availability: Host gene expression data are available under NCBI GEO accession number GSE168019 for bulk RNA-seq and GSE168018 for scRNA-seq. Raw microbial sequencing alignments are available from NCBI SRA under BioProject PRJNA704082. Code used for RNA-seq and microbiome analyses is available at https://github.com/bspeco/VAPinCOVID19.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.23.21253487; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

**Table 1**: Clinical characteristics of patients with COVID-19 who do or do not develop VAP at the "early" and "late" timepoints. Additional clinical and demographic characteristics can be found in Supplementary Tables 1 (bulk RNA-seq) and 2 (single cell RNA-seq).

| | VAP (early timepoint) | No-VAP (early timepoint) | P-value (early - VAP vs No-VAP) | VAP (late timepoint) | No-VAP (late timepoint) | P-value (late - VAP vs No-VAP) | Controls (early timepoint) | P-value (+/- COVID-19) |
|---|---|---|---|---|---|---|---|---|
| **Bulk RNA-seq** | | | | | | | | |
| N | 4 | 8 | | 5 | 8 | | 8 | |
| Age, median (IQR) | 59.5 (56.0, 66.0) | 49 (40.5, 64.0) | 0.268 | 51.0 (48.0, 53.0) | 59.5 (47.0, 64.5) | 0.271 | 63.0 (36.0, 79.0) | 0.642 |
| Female (%) | 1 (25.0%) | 3 (37.5%) | 1.00 | 2 (40.0%) | 4 (50.0%) | 1.00 | 4 (50.0%) | 0.648 |
| Days before VAP (median [range]) | 16.5 (15, 38) | - | - | 2 (2, 3) | - | - | - | - |
| Days post intubation (median [range]) | 2 (1, 5) | 2 (1, 4) | 0.930 | 17 (8, 23) | 9 (6, 19) | 0.160 | 1 (1, 3) | 0.062 |
| Cultured microbe in VAP patients (n) | | | | | | | | |
| *Klebsiella aerogenes* | 2 | - | - | 2 | - | - | - | - |
| *Staphylococcus aureus* | 1 | - | - | 0 | - | - | - | - |
| *Proteus mirabilis* | 1 | - | - | 1 | - | - | - | - |
| *Citrobacter koserii* | 0 | - | - | 1 | - | - | - | - |
| *Serratia marcescens* | 0 | - | - | 1 | - | - | - | - |
| *Pseudomonas aeruginosa* | 0 | - | - | 0 | - | - | - | - |
| **Single cell RNA-seq** | | | | | | | | |
| N | 4 | 6 | | 4 | 4 | | - | - |
| Age, median (IQR) | 64.0 (50.5, 76.5) | 56.5 (47.0, 68.0) | 0.392 | 61.5 (37.5, 74.0) | 60.0 (50.5, 74.0) | 0.271 | - | - |
| Female (%) | 2 (50.0%) | 2 (33.3%) | 1.00 | 0 (0.0%) | 2 (50.0%) | 0.429 | - | - |
| Days before VAP, median (range) | 9.5 (8, 16) | - | - | 3 (2, 4) | - | - | - | - |
| Days after intubation, median (range) | 5.5 (1, 10) | 4 (2, 13) | 0.748 | 8.5 (7, 18) | 10.5 (4, 15) | 0.884 | - | - |
| Cultured microbe in VAP patients (n) | | | | | | | | - |
| *Staphylococcus aureus* | 1 | - | - | 0 | - | - | - | - |
| *Klebsiella aerogenes* | 1 | - | - | 1 | - | - | - | - |
| *Citrobacter koserii* | 1 | - | - | 1 | - | - | - | - |
| *Proteus mirabilis* | 0 | - | - | 0 | - | - | - | - |
| *Serratia marcescens* | 0 | - | - | 1 | - | - | - | - |
| *Pseudomonas aeruginosa* | 1 | - | - | 1 | - | - | - | - |

Statistical significance was determined using Fisher's exact test (categorical variables) or by Wilcoxon test (continuous variables).
Abbreviations: IQR: interquartile range; VAP: ventilator-associated pneumonia.

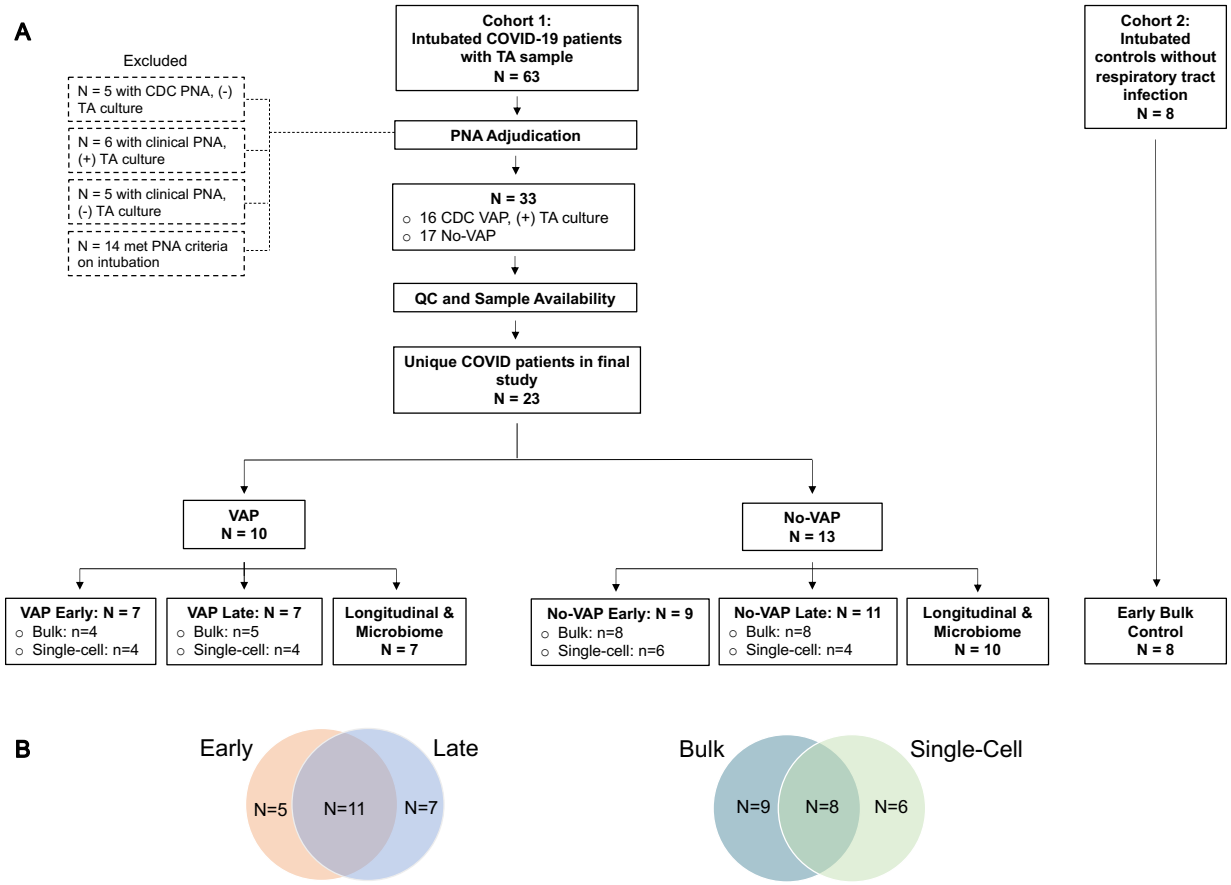



**Figure 1: Study flowchart.**
**A)** Two patient cohorts were studied. Cohort 1 consisted of COVID-19 patients from the COVID Multiphenotyping for Effective Therapies (COMET) / Immunophenotyping Assessment in a COVID-19 Cohort (IMPACC) studies (described in Methods). Cohort 2 consisted of critically ill intubated control patients from a prior prospective cohort study led by our research group[22]. The "early" samples were the first available tracheal aspirate specimens after intubation, and for the VAP group were collected >7 days prior to VAP onset. The "late" samples were obtained 2-4 days before VAP onset for the VAP group; for the No-VAP group, samples were selected to match the VAP group's time from intubation as closely as possible. Controls included eight critically ill, mechanically ventilated patients without LRTI. All COVID-19 patients included in the primary bulk analysis were also included in the longitudinal host expression and microbiome analyses. Patients who met CDC VAP criteria with negative TA culture, patients who met clinical VAP criteria, and patients with pneumonia upon intubation were excluded. See Supplementary Data File 8 for more details. **B)** The Venn diagram on the left illustrates that patients were included at the "early" timepoint, the "late" timepoint, or both, depending on sample availability. The Venn diagram on the right shows that subjects were included in the bulk sequencing analysis, the single-cell analysis, or both, again dependent on sample availability. Abbreviations: VAP=ventilator-associated pneumonia; TA=tracheal aspirate; QC=quality control; PNA=pneumonia; CDC=United States Centers for Disease Control and Prevention.

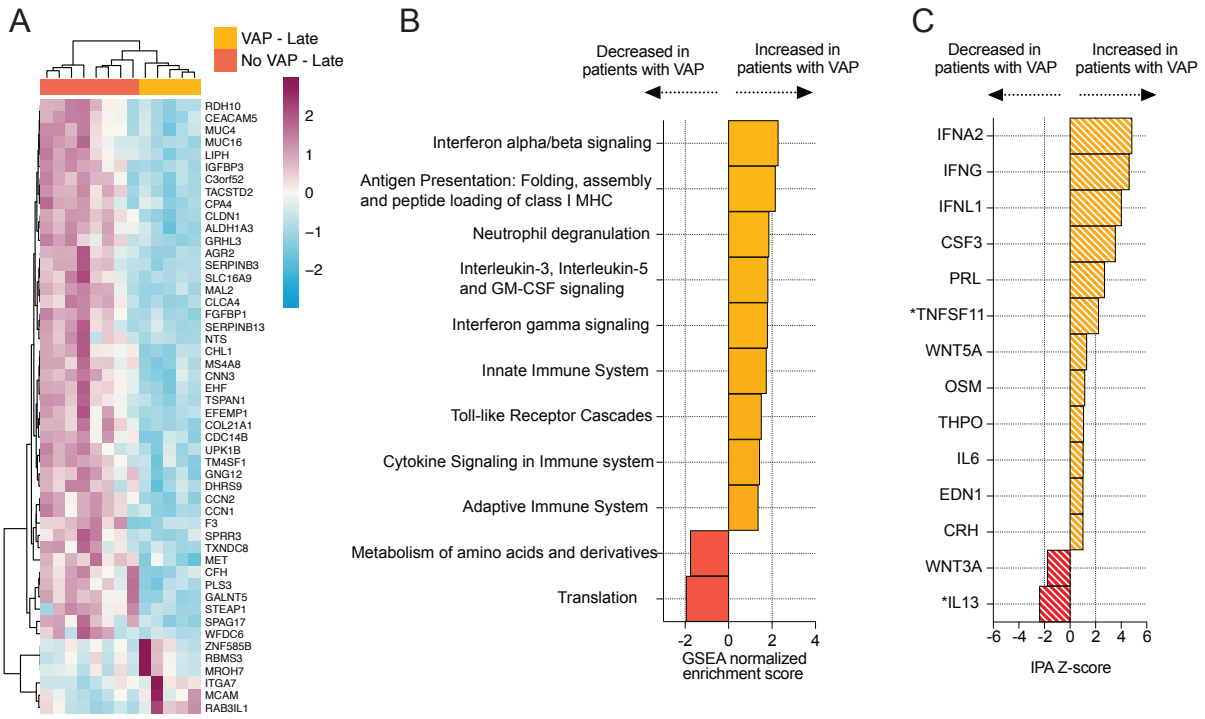
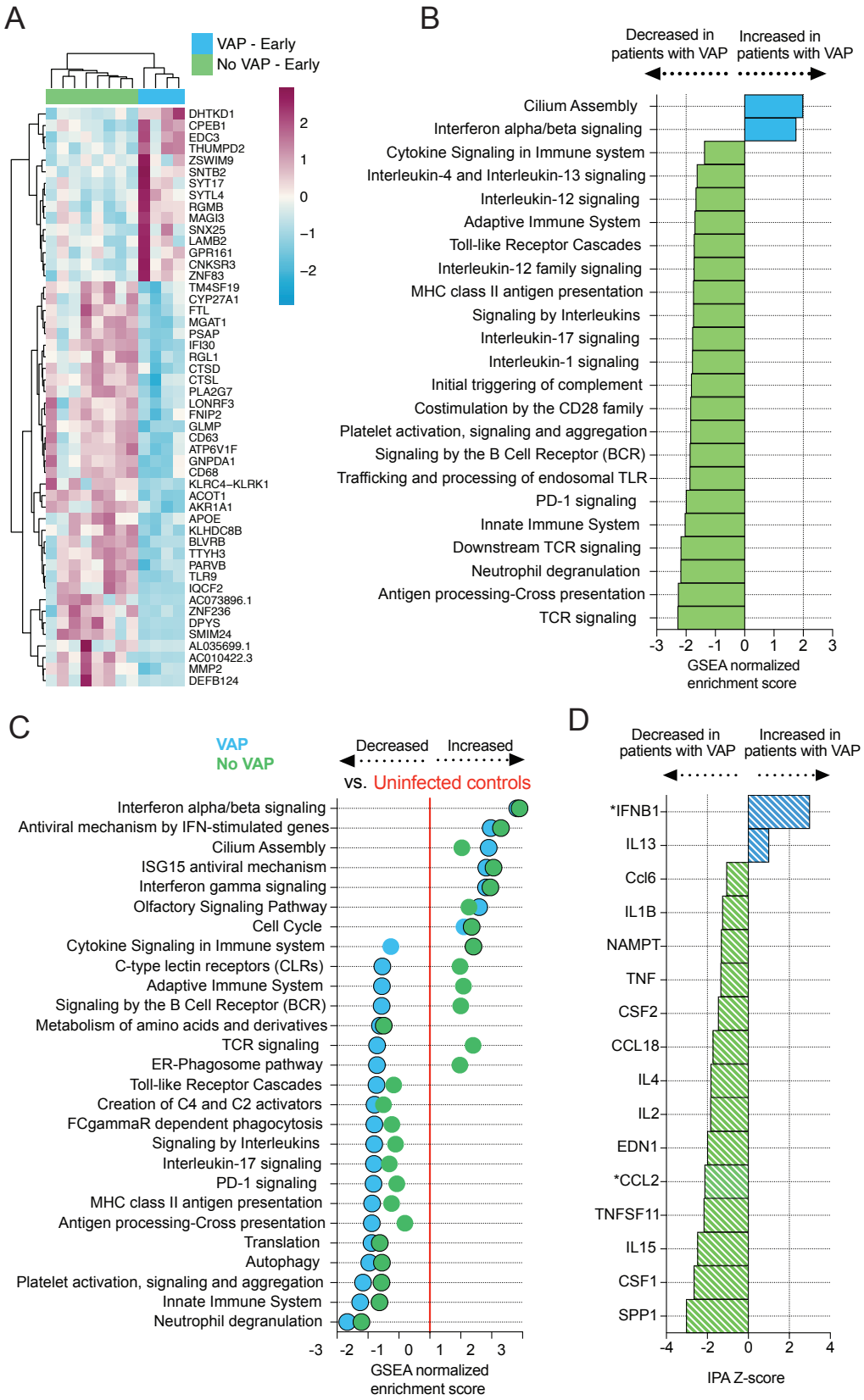
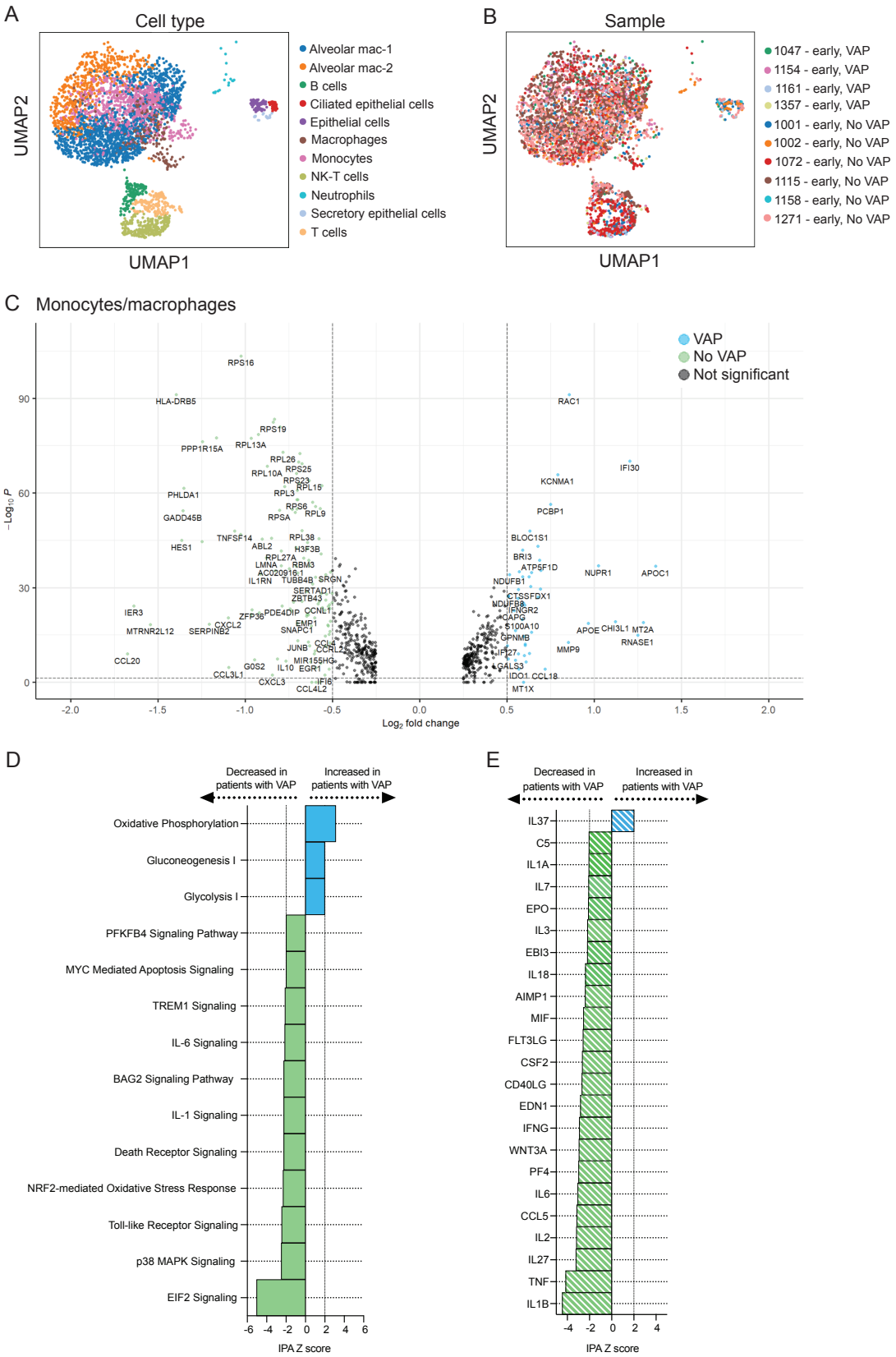



**Figure 2: COVID-19 VAP is associated with a lower respiratory tract transcriptional signature of bacterial infection 2 days before VAP onset.**
**A)** Heatmap of the top 50 differentially expressed genes by adjusted P-value between COVID-19 patients who developed VAP (yellow) versus those who did not (red) at the “late” timepoint, two days before the onset of VAP, from bulk RNA-seq. **B)** Gene set enrichment analysis (GSEA) at the “late” timepoint based on differential gene expression analyses. GSEA results were considered significant with an adjusted P-value < 0.05. **C)** Ingenuity Pathway Analysis (IPA) of upstream cytokines at the “late” timepoint based on differential gene expression analyses. IPA results were considered significant with a Z-score absolute value > 2 and overlap P-value < 0.05. *Denotes cytokines that were included with an overlap P-value < 0.1. All pathways and cytokines are shown in Supplementary Data Files 2 and 3.


A
VAP - Early
No VAP - Early
DHTKD1
CPEB1
EDC3
THUMPD2
ZSWIM9
SNTB2
SYT17
SYTL4
RGMB
MAGI3
SNX25
LAMB2
GPR161
CNKSR3
ZNF83
TM4SF19
CYP27A1
FTL
MGAT1
PSAP
IFI30
RGL1
CTSD
CTSL
PLA2G7
LONRF3
FNIP2
GLMP
CD63
ATP6V1F
GNPDA1
CD68
KLRC4–KLRK1
ACOT1
AKR1A1
APOE
KLHDC8B
BLVRB
TTYH3
PARVB
TLR9
IQCF2
AC073896.1
ZNF236
DPYS
SMIM24
AL035699.1
AC010422.3
MMP2
DEFB124
2
1
0
−1
−2
B
Decreased in patients with VAP
Increased in patients with VAP
Cilium Assembly
Interferon alpha/beta signaling
Cytokine Signaling in Immune system
Interleukin-4 and Interleukin-13 signaling
Interleukin-12 signaling
Adaptive Immune System
Toll-like Receptor Cascades
Interleukin-12 family signaling
MHC class II antigen presentation
Signaling by Interleukins
Interleukin-17 signaling
Interleukin-1 signaling
Initial triggering of complement
Costimulation by the CD28 family
Platelet activation, signaling and aggregation
Signaling by the B Cell Receptor (BCR)
Trafficking and processing of endosomal TLR
PD-1 signaling
Innate Immune System
Downstream TCR signaling
Neutrophil degranulation
Antigen processing-Cross presentation
TCR signaling
-3 -2 -1 0 1 2 3
GSEA normalized enrichment score
C
VAP
No VAP
Decreased
Increased
vs. Uninfected controls
Interferon alpha/beta signaling
Antiviral mechanism by IFN-stimulated genes
Cilium Assembly
ISG15 antiviral mechanism
Interferon gamma signaling
Olfactory Signaling Pathway
Cell Cycle
Cytokine Signaling in Immune system
C-type lectin receptors (CLRs)
Adaptive Immune System
Signaling by the B Cell Receptor (BCR)
Metabolism of amino acids and derivatives
TCR signaling
ER-Phagosome pathway
Toll-like Receptor Cascades
Creation of C4 and C2 activators
FCgammaR dependent phagocytosis
Signaling by Interleukins
Interleukin-17 signaling
PD-1 signaling
MHC class II antigen presentation
Antigen processing-Cross presentation
Translation
Autophagy
Platelet activation, signaling and aggregation
Innate Immune System
Neutrophil degranulation
-3 -2 -1 0 1 2 3
GSEA normalized enrichment score
D
Decreased in patients with VAP
Increased in patients with VAP
*IFNB1
IL13
Ccl6
IL1B
NAMPT
TNF
CSF2
CCL18
IL4
IL2
EDN1
*CCL2
TNFSF11
IL15
CSF1
SPP1
-4 -2 0 2 4
IPA Z-score

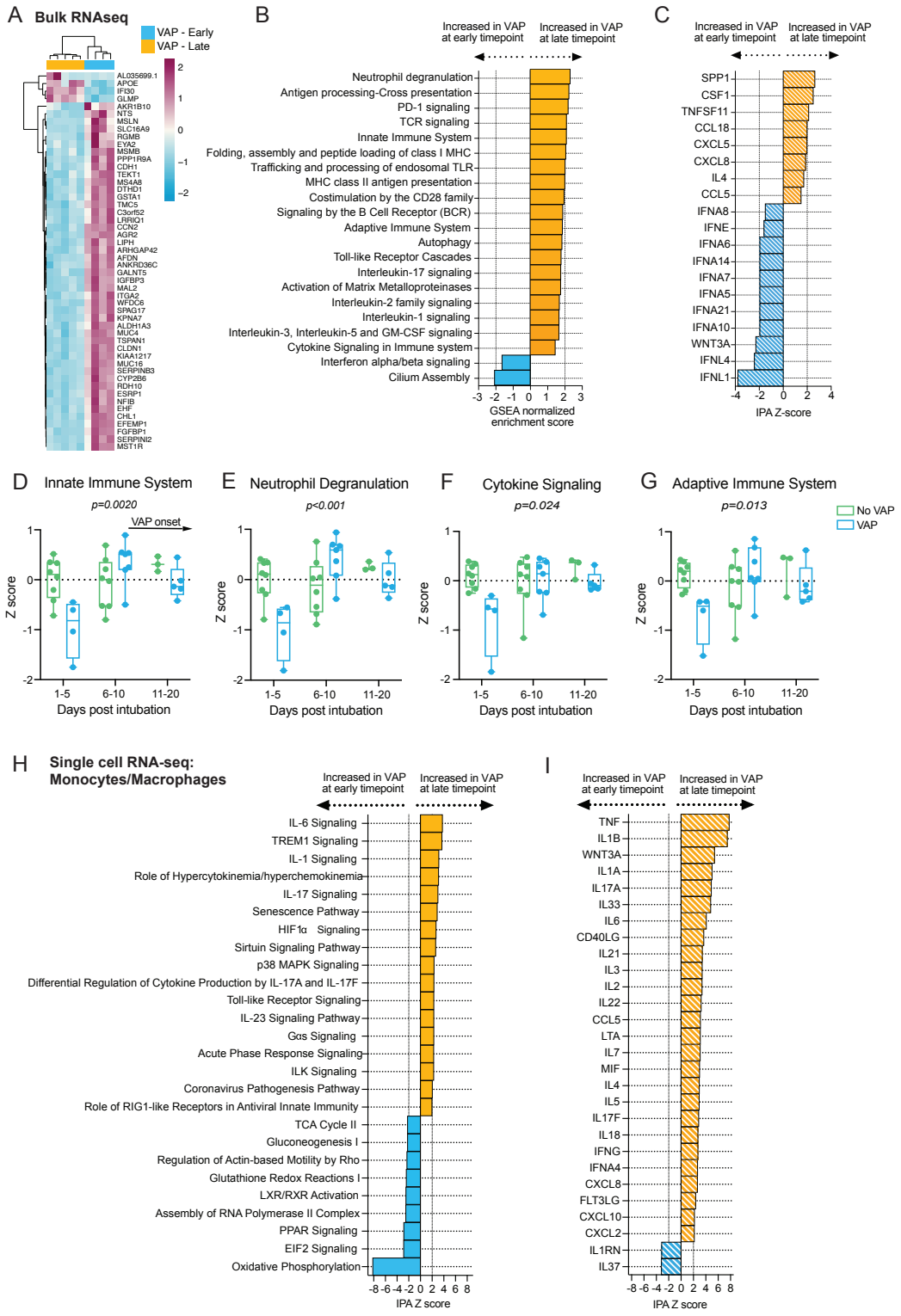
755

**Figure 3: COVID-19 patients who develop VAP have attenuated immune signaling in the lower respiratory tract two weeks before onset of secondary bacterial pneumonia.**
**A)** Heatmap of the top 50 differentially expressed genes by adjusted P-value between COVID-19 patients who developed VAP (blue) versus those who did not (green) at the "early" timepoint from bulk RNA-seq. **B)** Gene set enrichment analysis at the "early" timepoint based on differential gene expression analyses. GSEA results were considered significant with an adjusted P-value < 0.05. **C)** Expression of GSEA pathways at the "early" timepoint with respect to a baseline of uninfected, intubated controls. Pathways were selected from the GSEA results if they had an adjusted P-value < 0.05 in at least one of the comparisons (VAP vs controls or No-VAP vs controls). Pathways with an adjusted P-value < 0.05 when compared to controls are indicated by circles with a black outline. **D)** Ingenuity Pathway Analysis (IPA) of upstream cytokine activation states at the "early" timepoint based on differential gene expression analyses. IPA results were considered significant with a Z-score absolute value > 2 and overlap P-value < 0.05. *Denotes cytokines that were included with an overlap P-value < 0.1. All pathways and cytokines are shown in Supplementary Data Files 2 and 3.


A
Cell type
UMAP2
UMAP1
Alveolar mac-1
Alveolar mac-2
B cells
Ciliated epithelial cells
Epithelial cells
Macrophages
Monocytes
NK-T cells
Neutrophils
Secretory epithelial cells
T cells
B
Sample
1047 - early, VAP
1154 - early, VAP
1161 - early, VAP
1357 - early, VAP
1001 - early, No VAP
1002 - early, No VAP
1072 - early, No VAP
1115 - early, No VAP
1158 - early, No VAP
1271 - early, No VAP
C Monocytes/macrophages
VAP
No VAP
Not significant
−Log10 P
Log2 fold change
D
Decreased in patients with VAP
Increased in patients with VAP
Oxidative Phosphorylation
Gluconeogenesis I
Glycolysis I
PFKFB4 Signaling Pathway
MYC Mediated Apoptosis Signaling
TREM1 Signaling
IL-6 Signaling
BAG2 Signaling Pathway
IL-1 Signaling
Death Receptor Signaling
NRF2-mediated Oxidative Stress Response
Toll-like Receptor Signaling
p38 MAPK Signaling
EIF2 Signaling
IPA Z score
E
Decreased in patients with VAP
Increased in patients with VAP
IL37
C5
IL1A
IL7
EPO
IL3
EBI3
IL18
AIMP1
MIF
FLT3LG
CSF2
CD40LG
EDN1
IFNG
WNT3A
PF4
IL6
CCL5
IL2
IL27
TNF
IL1B
IPA Z score

771

772 **Figure 4: scRNA-seq demonstrates that COVID-19 VAP is associated with early impaired**
773 **antibacterial immune signaling in lower respiratory tract monocytes and macrophages.**
774 **A-B)** UMAP of single cell RNA-seq data from patients that do or do not develop VAP at the "early"
775 timepoint, annotated by (A) cell type or (B) sample. **C)** Volcano plot displaying the differentially
776 expressed genes between VAP and No-VAP patients in the low-resolution
777 monocytes/macrophages cluster. **D-E)** Ingenuity Pathway Analysis (IPA) of (D) key canonical
778 pathways and (E) upstream cytokines based on differential gene expression analysis in
779 monocytes/macrophages of COVID-19 patients who developed VAP versus those who did not.
780 Only significant pathways (IPA Z-score absolute value > 2 and overlap P-value < 0.05) are
781 depicted. All pathways and cytokines are provided in Supplementary Data Files 5 and 6.


A Bulk RNAseq
VAP - Early
VAP - Late
B
Increased in VAP at early timepoint
Increased in VAP at late timepoint
Neutrophil degranulation
Antigen processing-Cross presentation
PD-1 signaling
TCR signaling
Innate Immune System
Folding, assembly and peptide loading of class I MHC
Trafficking and processing of endosomal TLR
MHC class II antigen presentation
Costimulation by the CD28 family
Signaling by the B Cell Receptor (BCR)
Adaptive Immune System
Autophagy
Toll-like Receptor Cascades
Interleukin-17 signaling
Activation of Matrix Metalloproteinases
Interleukin-2 family signaling
Interleukin-1 signaling
Interleukin-3, Interleukin-5 and GM-CSF signaling
Cytokine Signaling in Immune system
Interferon alpha/beta signaling
Cilium Assembly
GSEA normalized enrichment score
C
SPP1
CSF1
TNFSF11
CCL18
CXCL5
CXCL8
IL4
CCL5
IFNA8
IFNE
IFNA6
IFNA14
IFNA7
IFNA5
IFNA21
IFNA10
WNT3A
IFNL4
IFNL1
IPA Z-score
D Innate Immune System
p=0.0020
VAP onset
E Neutrophil Degranulation
p<0.001
F Cytokine Signaling
p=0.024
G Adaptive Immune System
p=0.013
No VAP
VAP
Z score
Days post intubation
1-5
6-10
11-20
H Single cell RNA-seq: Monocytes/Macrophages
IL-6 Signaling
TREM1 Signaling
IL-1 Signaling
Role of Hypercytokinemia/hyperchemokinemia
IL-17 Signaling
Senescence Pathway
HIF1α Signaling
Sirtuin Signaling Pathway
p38 MAPK Signaling
Differential Regulation of Cytokine Production by IL-17A and IL-17F
Toll-like Receptor Signaling
IL-23 Signaling Pathway
Gαs Signaling
Acute Phase Response Signaling
ILK Signaling
Coronavirus Pathogenesis Pathway
Role of RIG1-like Receptors in Antiviral Innate Immunity
TCA Cycle II
Gluconeogenesis I
Regulation of Actin-based Motility by Rho
Glutathione Redox Reactions I
LXR/RXR Activation
Assembly of RNA Polymerase II Complex
PPAR Signaling
EIF2 Signaling
Oxidative Phosphorylation
IPA Z score
I
TNF
IL1B
WNT3A
IL1A
IL17A
IL33
IL6
CD40LG
IL21
IL3
IL2
IL22
CCL5
LTA
IL7
MIF
IL4
IL5
IL17F
IL18
IFNG
IFNA4
CXCL8
FLT3LG
CXCL10
CXCL2
IL1RN
IL37

782

783 **Figure 5: Temporal dynamics of the host response to VAP**
784 **A)** Heatmap of the top 50 differentially expressed genes by adjusted P-value between COVID-19
785 patients who developed VAP at the "early" timepoint (blue) versus the "late" timepoint (yellow)
786 from bulk RNA-seq. **B)** Gene set enrichment analysis (GSEA) based on differential gene
787 expression of VAP patients at the "early" vs "late" timepoint from bulk RNA-seq. GSEA results
788 were considered significant with an adjusted P-value < 0.05. **C)** Ingenuity Pathway Analysis (IPA)
789 of upstream cytokine activation states based on differential gene expression analyses of VAP
790 patients at the "early" vs "late" timepoint from bulk RNA-seq. IPA results were considered
791 significant with a Z-score absolute value > 2 and overlap P-value < 0.05. **D-G)** Longitudinal
792 analysis of selected pathway expression in VAP (blue) versus No-VAP (green) patients from bulk
793 RNA-seq samples taken from time of intubation to onset of VAP for all patients. Pathway Z-scores
794 were calculated by averaging Z-scores for the top 20 leading edge genes of each pathway,
795 determined by the results of GSEA comparing VAP versus No-VAP patients at the "early"
796 timepoint. Multiple Z-scores per patient at a given time interval were averaged so that each patient
797 corresponds to one datapoint at each interval. Samples from day 21+ after intubation are not
798 depicted due to a lack of these later timepoints in the No-VAP group. VAP onset in these patients
799 ranged from 10-39 days post intubation (median: 19 days). Selected pathways are: innate
800 immune system (D), neutrophil degranulation (E), cytokine signaling (F), and adaptive immune
801 system (G). Box plots represent the median and range. Statistical significance was determined
802 by two-way ANOVA, and interaction P-values are highlighted. **H-I)** Ingenuity Pathway Analysis
803 (IPA) of (H) key canonical pathways (I) and upstream cytokine activation states based on
804 differential gene expression analysis in monocytes/macrophages from scRNA-seq of COVID-19
805 patients who developed VAP versus those who did not. Only significant pathways (IPA Z-score
806 absolute value > 2 and overlap P-value < 0.05) are depicted. All pathways and cytokines are
807 provided in Supplementary Data Files 2, 3, 5, and 6.

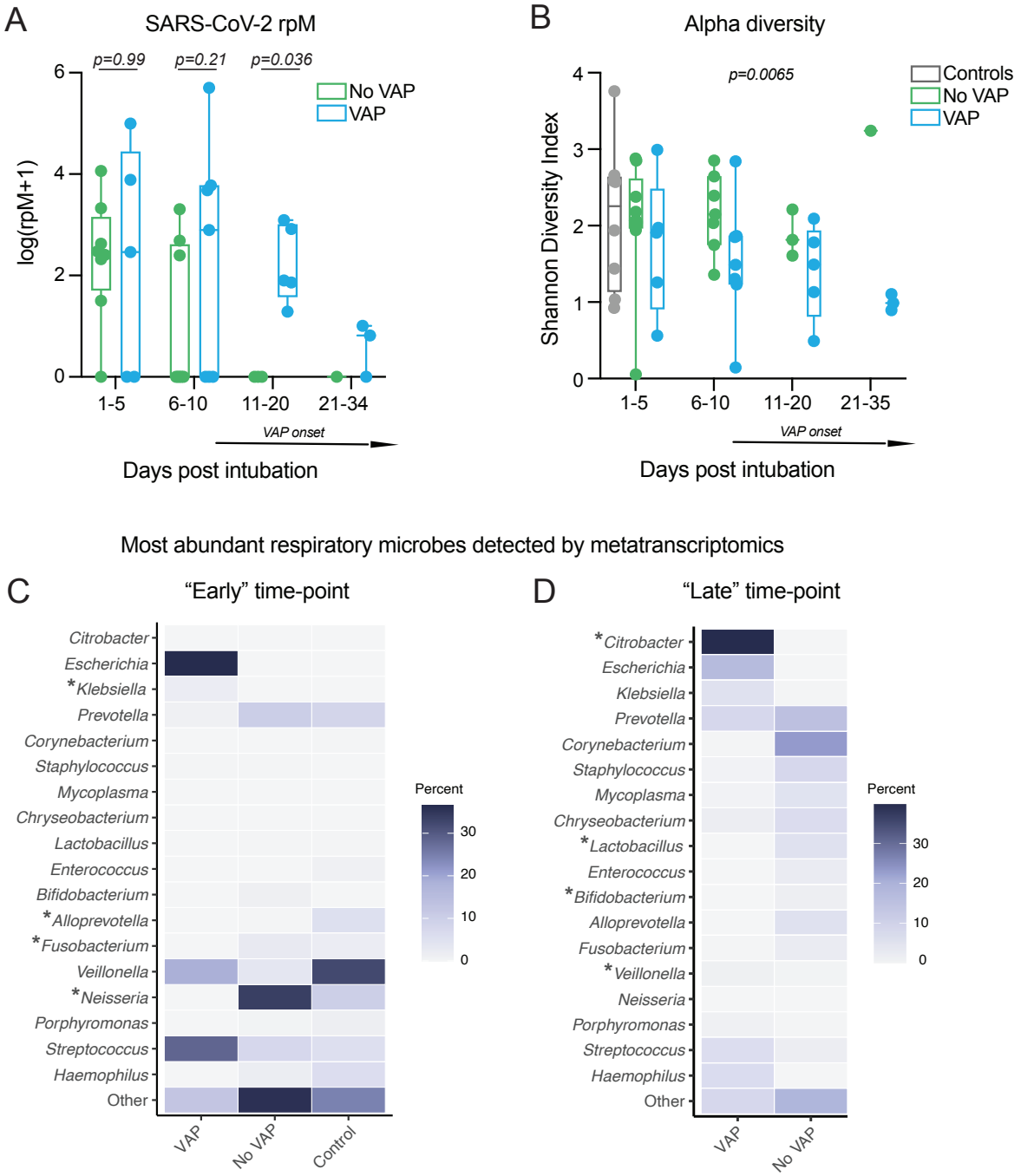



**Figure 6: Lung microbiome community collapse precedes VAP in COVID-19 patients.** **(A)** SARS-CoV-2 viral load (reads per million sequenced, rpM) over time by days since intubation in patients who develop VAP vs those who do not. For plotting purposes, log(rpM+1) was used to avoid negative values. **(B)** Lung microbiome bacterial diversity (Shannon's Index) in COVID-19 patients with relation to VAP development over time by days since intubation. VAP onset in these patients ranged from 10-39 days post intubation (median: 19 days). SARS-CoV-2 negative

816 intubated controls are included for comparison. Box plots represent the median and range.
817 Statistical significance was determined by Mann-Whitney tests (A) or two-way ANOVA between
818 VAP and No-VAP patients (B). P-values < 0.05 were considered significant. **(C-D)** Most abundant
819 respiratory microbes detected by metatranscriptomics at the (C) “early” and (D) “late” timepoints.
820 The mean abundance of the top 10 most abundant microbes from each group is depicted, and
821 significantly enriched taxa in VAP patients compared to No-VAP patients are marked by an
822 asterisk. SARS-CoV-2 negative intubated controls are included for comparison at the “early”
823 timepoint. Microbe abundance data for each patient can be found in Supplementary Data File 7.

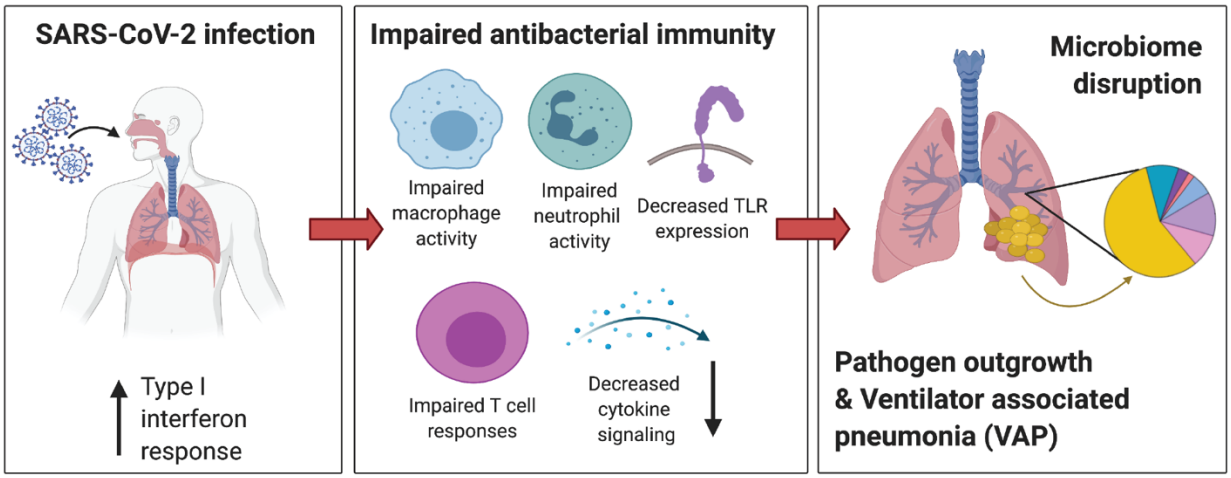



**Figure 7: Mechanistic hypothesis of secondary bacterial pneumonia susceptibility in patients with COVID-19.**
We hypothesize that individual immune responses to SARS-CoV-2 infection may drive a restructuring of the microbial community and increase susceptibility to VAP. Those predisposed to VAP have increased type I interferon responses and dysregulated antibacterial immune signaling characterized by impaired macrophage, neutrophil and T cell activity, decreased toll-like receptor (TLR) signaling and impaired activation of cytokines important for pathogen defense. This state of suppressed immunity may disrupt the lower respiratory tract microbiome, predisposing certain patients to the outgrowth of bacterial pathogens and VAP. Created with BioRender.com.