medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Mechanical ventilation affects respiratory microbiome of COVID-19 patients and its interactions with the host

Verónica Lloréns-Rico[1,2], Ann C. Gregory[1,2], Johan Van Weyenbergh[3], Sander Jansen[4], Tina Van Buyten[4], Junbin Qian[5,6], Marcos Braz[3], Soraya Maria Menezes[3], Pierre Van Mol[5,6,7], Lore Vanderbeke[8], Christophe Dooms[7,9], Jan Gunst[10], Greet Hermans[10], Philippe Meersseman[11], CONTAGIOUS collaborators, Els Wauters[7,9], Johan Neyts[4], Diether Lambrechts[5,6], Joost Wauters[11,12], Jeroen Raes[1,2,12,13]

[1] Laboratory of Molecular Bacteriology, Department of Microbiology and Immunology, Rega Institute, KU Leuven, Belgium

[2] Center for Microbiology, VIB, Leuven, Belgium

[3] Laboratory for Clinical and Evolutionary Virology, Department of Microbiology and Immunology, Rega Institute, KU Leuven, Belgium

[4] Laboratory of Virology and Chemotherapy, Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Belgium

[5] Laboratory of Translational Genetics, Department of Human Genetics, KU Leuven, Belgium

[6] VIB Center for Cancer Biology, VIB, Leuven, Belgium

[7] Department of Pneumology, University Hospitals Leuven, Belgium

[8] Laboratory of Clinical Bacteriology and Mycology, Department of Microbiology, Immunology and Transplantation, KU Leuven, Belgium

[9] Laboratory of Respiratory Diseases and Thoracic Surgery (BREATHE), Department of Chronic Diseases and Metabolism, KU Leuven, Belgium

[10] Laboratory of Intensive Care Medicine, Department of Cellular and Molecular Medicine, KU Leuven, Belgium

[11] Laboratory for Clinical Infectious and Inflammatory Disorders, Department of Microbiology, Immunology and Transplantation, KU Leuven, Belgium

[12] These authors contributed equally

[13] Corresponding author: jeroen.raes@kuleuven.vib.be

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2020.12.23.20248425; this version posted February 7, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Keywords: COVID-19, SARS-CoV-2, respiratory microbiome, single-cell RNA-sequencing, host-microbiome interactions

Abstract

Understanding the pathology of COVID-19 is a global research priority. Early evidence suggests that the microbiome may be playing a role in disease progression, yet current studies report contradictory results. Here, we examine potential confounders in COVID-19 microbiome studies by analyzing the upper (n=58) and lower (n=35) respiratory tract microbiome in well-phenotyped COVID-19 patients and controls combining microbiome sequencing, viral load determination, and immunoprofiling. We found that time in the intensive care unit and the type of oxygen support explained the most variation within the upper respiratory tract microbiome, dwarfing (non-significant) effects from viral load, disease severity, and immune status. Specifically, mechanical ventilation was linked to altered community structure, lower species- and higher strain-level diversity, and significant shifts in oral taxa previously associated with COVID-19. Single-cell transcriptomic analysis of the lower respiratory tract of ventilated COVID-19 patients identified increased oral microbiota compared to controls. These oral microbiota were found physically associated with proinflammatory immune cells, which showed higher levels of inflammatory markers. Overall, our findings suggest confounders are driving contradictory results in current COVID-19 microbiome studies and careful attention needs to be paid to ICU stay and type of oxygen support, as bacteria favored in these conditions may contribute to the inflammatory phenotypes observed in severe COVID-19 patients.

Introduction

COVID-19, a novel coronavirus disease classified as a pandemic by the World Health Organization, has caused over 40 million reported cases and 1 million deaths worldwide to date. Infection by its causative agent, the novel coronavirus SARS-CoV-2, results in a wide range of clinical manifestations: it is estimated that around 80% of

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

infected individuals are asymptomatic or present only mild respiratory and/or gastrointestinal symptoms, while the remaining 20% develop acute respiratory distress syndrome requiring hospitalization and oxygen support and, of those, 25% of cases necessitate critical care. Despite a concerted global research effort, many questions remain about the full spectrum of the disease severity. Independent studies from different countries, however, agree that age and sex are the major risk factors for disease severity and patient death[1–3], as well as type 2 diabetes and obesity [4,5]. Other risk factors for critical condition and death are viral load of the patient upon hospital admission[6–8] and the specific immune response to infection, with manifestation of a "cytokine storm" in critical patients characterized by increased levels of pro-inflammatory cytokines and chemokines, sustaining a disproportionate immune response that may ultimately cause organ failure[9–11].

Despite its close interplay with the immune system and its known associations with host health, little is known about the role of the respiratory microbiota in modulating COVID-19 disease severity, or its potential as a prognostic marker[12]. Previous studies exploring other pulmonary disorders have shown that lung microbiota may exacerbate their symptoms and contribute to their severity[13], potentially through direct crosstalk with the immune system and/or due to bacteremia and secondary infections[14]. Few studies of the respiratory microbiome in COVID-19 have revealed elevated levels of opportunistic pathogenic bacteria[15–17]. However, reports on bacterial diversity are contradictory. While some studies report a low microbial diversity in COVID-19 patients[15,18] that rebounds following recovery[16], others show an increased diversity in the COVID-19 associated microbiota[17]. These conflicting results could be due to differences in sampling location (upper or lower respiratory tract), severity of the patients, disease stage, or other confounders. While these early findings already suggest that the lung microbiome could be exacerbating or mitigating COVID-19 progression, exact mechanisms are yet to be elucidated. Therefore, an urgent need exists for studies identifying and tackling confounders in order to discern true signals from noise.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

To identify potential associations between COVID-19 severity and evolution and the upper and lower respiratory tract microbiota, we used nasopharyngeal swabs and bronchoalveolar lavage (BAL) samples, respectively. For the upper respiratory tract, we longitudinally profiled the nasopharyngeal microbiome of 58 COVID-19 patients during intensive care unit (ICU) treatment and after discharge to a classical hospital ward following clinical improvement, in conjunction with viral load determination and nCounter immune profiling. For the lower respiratory tract, we analyzed microbial signals in cross-sectional single-cell RNA-seq data from of bronchoalveolar lavage (BAL) samples of 22 COVID-19 patients and 13 pneumonitis controls with negative COVID-19 qRT-PCR, obtained from the same hospital. The integration of these data enabled us to (1) identify potential confounders of COVID-19 microbiome associations, (2) explore how microbial diversity evolves throughout hospitalization, (3) study microbe-host cell interaction and (4) substantiate a link between the respiratory microbiome and SARS-CoV-2 viral load as well as COVID-19 disease severity. Altogether, our results directly point to specific interactions between the microbiota and the immune cells, likely driven by clinical ventilation practices, which could potentially influence COVID-19 disease progression and resolution.

Results

The upper respiratory microbiota of COVID-19 patients

We longitudinally profiled the upper respiratory microbiota of 58 patients diagnosed with COVID-19 based on a positive qRT-PCR test or a negative test with high clinical suspicion based on symptomatology and a chest CT-scan showing alveolar damage. All these patients were admitted and treated at UZ Leuven hospital. Patient demographics for this cohort are shown in Table 1.

In total, 112 nasopharyngeal swabs from these patients were processed (Figure 1a): the V4 region of the 16S rRNA gene amplified on extracted DNA using 515F and 806R primers, and sequenced on an Illumina MiSeq platform (see Methods). From the same swabs, RNA was extracted to determine SARS-CoV-2 viral loads and to estimate

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

immune cell populations of the host using nCounter (Methods). Of the 112 samples processed and sequenced, 101 yielded over 10,000 amplicon reads that could be assigned to bacteria at the genus level (Figure 1b; Methods). The microbiome of the entire cohort was dominated by the gram-positive genera *Staphylococcus* and *Corynebacterium*, typical from the nasal cavity and nasopharynx[19].

Bacterial alpha diversity is strongly associated with ICU stay length

First, we determined genus-level alpha-diversity for the 101 samples with more than 10,000 assigned reads, using Shannon Diversity index (see Methods; Supplementary Table 1). We observed that alpha diversity was not significantly correlated to SARS-CoV-2 viral load in nasopharyngeal swabs (Figure 1c). In contrast, we found the Shannon index to be significantly affected by the sampling moment (Kruskal-Wallis test, p-value = 0.0076; Figure 1d), with significant differences between swabs procured upon patient admission and later timepoints, suggesting an important effect of disease progression and/or treatment (i.e. due to antibiotics administered throughout ICU stay). We explored these differences further, and observed that Shannon Diversity index correlated negatively with the number of days spent in ICU at the moment of sampling, with longer ICU stays leading to a lower diversity (Supplementary Figure 1a; ρ=-0.55, p-value=$4.4 \cdot 10^{-9}$). Furthermore, the observed decrease in diversity occurred mostly over the first 2 weeks in ICU (Supplementary Figure 1a). Other severity indicators, such as the patient clinical status (i.e. a qualitative metric used to classify patients into different levels of disease severity) or the type of oxygen support required at the moment of sampling, showed no association with the genus-level Shannon index (Supplementary Figure 1b,c). Therefore, the differences in diversity observed across samples were likely driven by the time spent in ICU, and not specifically by disease progression as other severity indicators were unaffected.

Since time in ICU had a major effect in alpha diversity, we explored whether it might be masking any effects of viral load, as SARS-CoV-2 load was not associated with the length of the ICU stay (Supplementary Figure 1d). When controlling for the time in ICU (see Methods), we observed that viral load was negatively associated with the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

Shannon diversity index (ρ=-0.26, p-value=0.0089; Supplementary Figure 1e). Other severity indicators such as the clinical status and the type of oxygen support were not correlated to diversity even after controlling for the confounding effect of ICU stay (Kruskal-Wallis test, p-value>0.05).

Microbiome composition variation is driven by the type of respiratory support

We next explored potential associations between the upper respiratory genus-level microbiota composition and the extensive metadata collected in the study. In total, 70 covariates related to patient anthropometrics, medication and clinical variables measured in the hospital and host cytokine expression measured in the swabs were tested (Supplementary Table 2). Individually, 19 of these covariates showed a significant correlation to microbiota composition (dbRDA, p-value<0.05;FDR<0.05; Figure 2a). These significant covariates were related to disease and measures of its severity, such as the clinical evaluation of the patient, the total length of the ICU stay, the number of days in ICU at the time of sampling, or the type of oxygen support needed by the patient. Surprisingly, the total SARS-CoV-2 viral load detected in the swabs was not significantly associated to microbiome composition variation (Supplementary Table 2).

Of the 19 significant covariates, only 2 accounted for 48.7% non-redundant variation in this dataset, with the rest holding redundant information. These were the patient ID, included due to the longitudinal sampling of patients, and confirming that intra-individual variation over time is smaller than patient inter-individual variation[20], and the type of oxygen support received at the time of sampling (Figure 2a,b). Notably, the type of oxygen support discriminated samples based on ventilation type, with non-invasive ventilation samples (groups 1, 2 and 3) separating from samples from intubated patients (groups 4 to 7; PERMANOVA test, R^2=0.0642, p-value=0.001).

To determine if oxygen support also impacted the microbiome at finer taxonomic resolution, we revisited alpha-diversity at species- and strain-level. We defined species as 97% identity 16S OTUs and strains per species as the clustered 16S sequences

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

within each OTU. The species and strain-level diversity per sample were calculated as the number of OTUs and as the mean of the number of strains from five randomly sampled OTU species sampled 1,000 times, respectively. Our analyses revealed both species- and strain-level diversity change with ventilation, even with non-invasive ventilation (e.g. BIPAP, CPAP). Across all samples we observed high species- and low strain-level diversity pre-ventilation, which reversed following any form of ventilation (Figure 2c; Wilcoxon test; p-values<0.05, with the exception of type 7), with the exception of ventilation with inhaled nitric oxide. Further, Species- and strain-level diversity showed a strong inverse correlation (Figure 2d; Pearson's correlation, R^2 = -0.92, p-value = 0.0035).

Therefore, we evaluated which specific taxa were differentially abundant between samples from intubated and non-intubated patients. In total, 30 genera were more abundant in intubated samples, while 2 genera were more abundant in non-invasively ventilated patients (p-value<0.05;FDR<0.05; Figure 2e, Supplementary Figure 2; Supplementary Table 3). Some of these taxa are common oral microbiome commensals or opportunistic pathogens that had been repeatedly reported as more abundant in COVID-19 patients than in healthy controls, such as *Prevotella*, *Veillonella*, *Fusobacterium*, *Porphyromonas* or *Lactobacillus*[15–17]. Here, we reported higher abundance of these genera in intubated COVID-19 patients as compared to non-mechanically ventilated patients. This points at mechanical ventilation as a potential confounder of previous COVID-19 studies. Additionally, we found other taxa not previously reported in previous COVID-19 microbiome studies, such as *Mycoplasma* or *Megasphaera* (Figure 2e, Supplementary Figure 2), but previously associated to risk of ventilator-associated pneumonia[21].

By extracting the amplicon sequence variants (ASVs) corresponding to these differentially abundant genera (see Methods), some of these taxa could be narrowed down to the species level, confirming their origin as typically oral bacteria: for instance, *Prevotella* species included *P. oris*, *P. salivae*, *P. denticola*, *P. buccalis* and *P. oralis*. Within the *Mycoplasma* genus, ASVs were assigned to *Mycoplasma salivarium* among other species, an oral bacterium which has been previously associated to the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

incidence of ventilator-associated pneumonia[21]. When controlling for ventilation type, no taxa were found associated to SARS-CoV-2 viral loads. These results show that further research with larger cohorts and controlling for the relevant confounders highlighted here, such as ventilation type or length of stay in ICU, will be needed to study the specific effect of the viral infection.

Single-cell RNA-seq identifies oral commensals and opportunist pathogens in the lower respiratory tract

Next, we explored what the functional consequences of (ventilation-driven) lung microbiome disturbances could be. To do so, we screened single-cell RNA-seq data generated on BAL samples of 35 patients[22] to identify microbial reads. All patients in this cross-sectional cohort had clinical symptoms of pneumonia, 22 of them being diagnosed with COVID-19. The other 13 patients with non-COVID-19 pneumonia were hereafter referred to as controls (Table 1). Microbiome read screening of these samples revealed an average of 7,295.3 microbial reads per sample (ranging from 0 to 74,226 reads, with only a single sample yielding zero microbial reads; Supplementary Figure 3).

Among the top taxa encountered in these patients, we found some similarities with the data obtained in nasopharyngeal swabs. The top 15 species detected include *Mycoplasma salivarium* as the dominating taxon in 5 COVID-19 patients in ICU, as well as different *Prevotella* members. Non-COVID-19 pneumonia patients in ward (i.e. non-intubated) harbored different microbes: 2 patients had a microbiome dominated by *Porphyromonas gingivalis*, while a single patient had a microbiome dominated by the fungus *Pneumocystis jirovecii*, a known pathogen causing pneumonia[23].
Supplementary table 4 shows associations between organism abundances and specific patient metadata: disease, hospital stay and ventilation type. Multiple links with COVID-19 diagnosis were identified (Wilcoxon test, (noncorrected) p-value<0.05; see Methods) but due to the low sample number, none was significant after multiple-test correction. Additionally, as hospital stay (ICU or ward), type of oxygen support (invasive or non-invasive ventilation) and disease (COVID-19 or controls) were highly

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

correlated (Chi-squared test, p-value < 0.0001 for all three pairwise correlations), the effect of these three variables could not be disentangled.

Bacteria in the lower respiratory tract associate to host cells from the innate immune system in COVID-19 patients

Next, we took advantage of the single-cell barcoding and questioned whether the microbial cells that we identified were found in association with host cells, or contrarily, had unique barcodes suggesting a free-living state. In total, 29,886 unique barcodes were identified that matched a total of 46,151 microbial UMIs. The distribution of UMIs per barcode was asymmetrical, ranging from 1 to 201 and with 88% of the barcodes having a single UMI. Additionally, 26,572 barcodes (89%) were associated to a single microbial species, the rest being associated to 2 species (8.8%) or more (2.2%).

Out of the total 29,886 microbial barcodes, only 2,108 were also assigned to host cells, suggesting that the bulk of bacteria found in BAL samples exist as free-living organisms or in bacterial biofilms. However, for those associated to host cells, the distribution across disease types was not random. We found that while 2.3% of the non-COVID-19 patient cells were associated to bacterial cells, almost the double (4%) could be observed in COVID-19 patients (Figure 3a; Chi-squared test; p-value < $2.2 \cdot 10^{-16}$). However, because COVID-19 diagnosis is highly correlated with intubation in this cohort, this effect could be due to higher intubation rates in COVID-19 patients. Within COVID-19 patients, we also evaluated the overlap between bacteria-associated host cells and cells with detected SARS-CoV-2 reads (Supplementary Table 5). Out of 1033 host cells associated with bacteria in these patients and 343 cells with detected SARS-CoV-2 reads, only one cell was positive for both. A binomial test for independence of virus and bacteria detection in the same host cell, showed that the observed co-occurrence in one cell only was highly unlikely (p-value=$5.7 \cdot 10^{-4}$), therefore suggesting mutual exclusion of microbiome members and viruses in the same host immune cells.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

We also explored whether host-associated bacterial reads would preferentially be linked with specific cell types, taking into account the varying frequencies of cell types in COVID-19 patients and controls (see Methods). Such a preferential association would suggest that these observations are biologically relevant and not an artifact of the single-cell sample and library preparation. Among control patients, cell types were similarly distributed in both groups (i.e. with and without bacteria), with only a preferential association of microbial cells with neutrophils (p-value = $3.61 \cdot 10^{-12}$; Figure 3b; Supplementary Figure 4). However, in COVID-19 patients, three cell types were significantly associated with bacteria: neutrophils (p-value < $2.2 \cdot 10^{-16}$), monocytes ((p-value = $4.82 \cdot 10^{-5}$) and monocyte-derived macrophages (p-value < $2.2 \cdot 10^{-16}$; Figure 3b; Supplementary Figure 4). We also found that different bacteria associate with distinct host cells. For instance, in COVID-19 patients, bacteria from the *Mycoplasma* genus preferentially associated to monocyte-derived macrophages (p-value = $2.28 \cdot 10^{-7}$), while *Rothia* (p-value = $8.21 \cdot 10^{-4}$), *Enterobacter* (p-value = $2.59 \cdot 10^{-5}$), or *Klebsiella* (p-value = $3.12 \cdot 10^{-9}$) are enriched in monocytes (Figure 3c).

Last, we investigated whether the associations of bacteria to host cells are linked to host cell expression. To do so, we assessed whether expression based cell subtype classification[22] for neutrophils, monocytes and macrophages showed non-random associations with bacteria across all samples in this cohort. Among the neutrophils, a subtype of inflammatory neutrophils characterized by expression of the calgranulin S100A12 was enriched in bacteria-associated cells (p-value $7.18 \cdot 10^{-6}$; Figure 3d,e). This subset of cells was also found to be enriched in SARS-CoV-2 nucleocapsid gene reads[22], suggesting that the same cell type responsible for defense against the virus would be responding to potentially invasive bacteria in the lung. This subgroup is characterized by the expression of the calprotectin subunits S100A8 and S100A9. It is known that S100A8/A9 heterodimer secretion is increased in infection-induced inflammation and has some antibacterial effects mediated by secretion of pro-inflammatory cytokines, release of reactive oxygen species and recruitment of other inflammatory cells, as well as chelation of Zn^{2+} necessary for bacterial enzymatic activity[24]. These mechanisms are mediated by binding of the S100A8/A9 dimer to TLR4 receptors to trigger the release of pro-inflammatory cytokines such as IL-6 and TNF-α, and thus may contribute to

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

sustain or exacerbate inflammation[25]. Therefore, the association with bacteria could, at least in part, explain the inflammatory phenotype of this neutrophil subset. To further examine this hypothesis, we explored differential gene expression between bacteria-associated and non-associated S100A12^{hi} neutrophils (Supplementary Table 6). Because association of these cells with SARS-CoV-2 and with bacteria was mutually exclusive, we also compared these changes with the ones triggered by the virus in neutrophils[22]. Within this subset, neutrophils with co-occurring bacteria showed significantly higher expression (Bonferroni-corrected p-value < 0.05) of pro-inflammatory genes, including the cytokine IL1B and some of its target genes (PTSG2), the transcription factors FOS and JUN, and several genes involved in degranulation (S100A9, FOLR3, HSPA1A, HSP90AA1, FCGR3B), (Supplementary Table 6). Among these, FOLR3, a gene encoding for a folate receptor, is found in neutrophil secretory granules and has antibacterial functions, by binding folates and thus depriving bacteria of these essential metabolites[26]. This response differed to that of virus-engulfing neutrophils in that IFN response genes are not distinctively upregulated by bacteria.

Regarding myeloid cells, both inflammatory IL1B^{hi} monocytes (p-value = $2 \cdot 10^{-16}$) as well as a mixed group of CCL2-expressing macrophages (p-value = $5.38 \cdot 10^{-10}$) are enriched in bacteria-associated cells (Figure 3f). These inflammatory monocytes are believed to have an important role in the cytokine storm occurring in severe COVID-19 patients. In this case, further gene expression patterns were detected, specific for bacteria-associated cells: for CCL2^{hi} macrophages, cells with co-occurring bacteria expressed higher levels of MHC genes of type I and II, suggesting a more active role of these cells in antigen presentation (Bonferroni-corrected p-value < 0.05; Figure 3f; Supplementary Table 6). A similar increase was also observed in monocytes, yet not significant (Supplementary Table 6), possibly due to the lower monocyte abundances in this dataset. Additionally, bacteria-associated macrophages express significantly higher levels of the calprotectin subunits S100A8/A9, similarly to neutrophils, as well as pro-inflammatory chemokines (such as CCL4, CXCL10 and CXCL1).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

Altogether, our results suggest that the bacteria detected in these cell subsets via scRNA-seq analyses may be contributing to the inflammatory response observed in the host.

Discussion

Since the beginning of the COVID-19 pandemic, a massive global effort by the scientific community was undertaken to understand physiopathology of SARS-CoV-2 infection and risk factors affecting disease outcome. In this study, we explored the respiratory microbiota as a potential risk factor for disease severity, and we evaluated the upper and lower respiratory tract microbiota in COVID-19 patients throughout their hospitalization. We linked this data to viral load measurements and immunoprofiling results from nCounter and single-cell RNA sequencing data. To assess robustness of previously reported signals, we investigated the effect of potential confounders based on a broad panel of patient metadata variables.

We found that in the upper respiratory tract, while SARS-CoV-2 viral load has a weak negative association with bacterial biodiversity, a strong effect of severity indicators such as ICU stay was observed, with diversity decreasing throughout the length of the ICU period, a pattern reminiscent of that seen in other pulmonary conditions[27,28].
This effect of ICU and/or ventilation on microbiome alpha diversity could potentially explain why previous studies on the microbiota of COVID-19 patients show conflicting results regarding diversity: some studies reported lower diversity in sputum or throat swab samples of COVID-19 patients[15,16,18] while others focusing on the lower respiratory microbiome using bronchoalveolar fluid samples, showed higher bacterial diversity in COVID-19 patients than in controls[17]. To further complicate matters, it cannot be excluded that sampling site or processing could also be potential confounders in these studies and/or reflect the different pathologies in the different areas of the respiratory tract.

We further found that between patient microbiome variation (as measured by genus-level microbial beta-diversity) was also affected by different severity indicators such as

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

the clinical status of the patient, or more importantly the type of oxygen support received, with intubated patients harboring a different microbiota than non-intubated patients. The impact of oxygen support was also reflected at the species- and strain-levels, with intubation causing a significant decrease and increase, respectively, in diversity. We hypothesize that the introduction of forced oxygen may drive the fast extinction of certain microbial species enabling the diversification of existing or newly colonizing species into new strains. Our results suggest that non-invasive ventilation (e.g. BIPAP, CPAP) can have microbial effects indicating that any form of ventilation may be a tipping point for microbial community differences.

Importantly, several of the taxa reported to change between intubated and non-intubated patients were reported to be linked to diagnosis in previous COVID-19 microbiome studies[15–17]. In our study, no taxa were specifically linked to SARS-CoV-2 viral load after controlling for intubation. This result strongly points at the possibility that intubation and mechanical ventilation are confounding previous results. Indeed, one study comparing COVID-19 patients with patients diagnosed of community-acquired pneumonia found no differences in respiratory microbiome composition between both groups of patients, but both groups did differ from healthy controls[29]. Together, these results indicate that patient intubation or even non-invasive ventilation are to be considered as important confounders when studying the upper respiratory microbiome, and we strongly suggest future COVID-19 microbiome studies should foresee and include strategies to account for this covariate. A recent study found a single ASV corresponding to the genus *Rothia* that was specific for SARS-CoV-2 patients after controlling for ICU-related confounders[30].

To better understand the potential functional consequences of these procedures and linked microbial shifts, we also profiled the microbiome of the lower respiratory tract using single-cell data obtained from a cross-sectional cohort of patients derived from the same hospital. Our results show that 'standard' single-cell RNA-seq, even though not optimized for microbial detection and profiling, can identify bacteria alone or in association with specific human cells. Unfortunately, the low numbers of microbial reads obtained in this cohort, together with the fact that ICU stay, COVID-19 diagnosis

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

and intubation are highly correlated in this set of patients, only allow for a descriptive analysis of the results. In this cohort, we identified different oral commensals and opportunistic pathogens previously linked to COVID-19 patients in both groups of samples, thus pointing again at a potential ventilation-linked origin. More interestingly, we identified a subset of bacteria associated with host cells, more specifically with neutrophils, monocytes and macrophages. This enrichment shows that these bacteria are likely not random contaminants, from which an even distribution across cell types (i.e. considering cell type abundances) would be expected. The identity of these host cells suggests that bacteria could have been phagocyted by these innate immune system cells, rather than be attached to the host cell surface. To the best of our knowledge, this is the first study linking interacting host cells and lung microbiome via high-throughput single-cell RNA-seq.

We find that host cells associated with bacteria, most of which are of oral origin, exhibit pro-inflammatory phenotypes as well as higher levels of MHC for antigen presentation. In this single-cell cohort it was observed that critical COVID-19 patients are characterized by an impaired monocyte to macrophage differentiation, resulting in an excess of pro-inflammatory monocytes, as well as by prolonged neutrophil inflammation[22]. Given that only these cell types are enriched in bacteria, we hypothesize that the respiratory (or ventilation-linked) microbiome may be playing a role in exacerbating COVID-19 progression in the lower respiratory tract. We verified that this response would likely be driven by bacteria and not SARS-CoV-2, which is also detected mostly in these cell types, as there is almost no overlap in detection of both virus and bacteria in the same cells. However, it must be noted that lack of detection does not completely rule out presence of virus and bacteria within these cells. Therefore, further research is required in order to confirm a causative role of the microbiota in this immune impairment characteristic of critical disease, and to reveal the specific mechanisms involved.

The presence of oral taxa in the lung microbiota is not unique of disease conditions. It is known that microaspiration, or the aspiration of aerosol droplets originated in the oral cavity, occurs in healthy individuals and can serve as a route for lung colonization

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

of oral commensals[31]. Such an increase of oral bacteria in the lower respiratory tract could be facilitated when critically ill patients –including but not limited to COVID-19– get intubated. Indeed, oral bacteria have been linked to ventilator-associated pneumonia[32,33]. It is yet to be elucidated whether COVID-19 physiopathology favors lung colonization by oral bacteria or if, in contrast, a lung microbiome previously colonized by oral microbes could also contribute to the disease. What is known is that an increase of oral bacteria in the lower respiratory tract can result in an increased inflammatory phenotype, even in healthy subjects[34]

Conclusion

Overall, this study provides a systematic analysis of potential confounders in COVID-19 microbiome studies. We identified that ICU hospitalization and type of oxygen support had large impacts on the upper respiratory tract microbiome and have the potential to confound clinical microbiome studies. Among the different types of oxygen support we reported the largest shifts in microbial community structure between intubated and non-intubated patients. We found that oral microbiota was strongly enriched in the upper and lower respiratory tracts of intubated COVID-19 patients. Further, in the lower respiratory tract, microbes were strongly associated with specific pro-inflammatory immune cells. This information contributes to a collective body of literature on the pathology of COVID-19 and suggests that careful attention be paid to ICU stay and type of oxygen support when evaluating the role of the lung microbiome on COVID-19 disease progression.

Methods

Study design and patient cohorts

All experimental protocols and data analyses were approved by the Ethics Commission from the UZ Leuven Hospital, under the COntAGIouS observational clinical trial (study number S63381). The study design is compliant with all relevant ethical regulations,

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

including the Declaration of Helsinki and in the GDPR. All participants gave their informed consent to participate in the study.

A total of 58 patients from the COntAGIouS observational trial were included as our upper respiratory tract cohort. All patients were admitted to the UZ Leuven hospital with a diagnostic of COVID-19. The disease was diagnosed based on a) a positive qRT-PCR test, performed on admission or previously on other hospitals, when patients were transferred from other medical facilities; or b) a chest CT-scan showing alveolar damage and clinical symptoms of the disease. All patients included in the study were admitted to ICU for a variable amount of time. Nasopharyngeal swabs were taken from these patients at different timepoints throughout ICU stay and after ICU discharge, during recovery in ward. A total of 112 swabs were processed for upper respiratory microbiome characterization (Figure 1a).

To extend our findings from the upper respiratory tract, we also profiled the lower respiratory tract microbiota in a different cohort[22] of 35 patients belonging to the same observational trial and also recruited at UZ Leuven hospital. This cross-sectional cohort is composed by 22 COVID-19 patients and 13 pneumonitis controls with negative qRT-PCR for SARS-CoV-2, with varying disease severity. Previous data from single-cell RNA-sequencing had been collected for this cohort[22]. We reanalyzed this single-cell dataset to profile the lower respiratory tract microbiota in these patients.

RNA/DNA extraction and sequencing

Nucleic acid extraction from the swab samples was performed with AllPrep DNA/RNA/miRNA Universal kit (QIAGEN, catnr. 80224). Briefly, swabs from the potentially infectious samples were inactivated by adding 600µL RLT-plus lysis buffer. To increase bacterial cell lysis efficiency, glass beads and DX reagent (Pathogen Lysis Tubes, QIAGEN, catnr. 19091) were added to the lysis buffer, and samples were disrupted in a FastPrep-24™ instrument with the following program: 1-minute beating at 6.5m/sec, 1-minute incubation at 4°C, 1-minute beating at 6.5m/sec, 1-minute

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

incubation at 4°C. After lysis, the remaining extraction steps followed the recommended protocol from the manufacturer. DNA was eluted in 50µL EB buffer. Amplification of the V4 region of the 16S gene was done with primers 515F and 806R, using single multiplex identifiers and adaptors as previously described[35]. RNA was eluted in 30µL of nuclease-free water and used for SARS-CoV-2 viral load determination in the swabs as well as to measure inflammatory markers and cytokines and to estimate host cell populations via marker gene expression using nCounter. In brief, raw nCounter data were processed using nSolver 4.0 software (Nanostring), sequentially correcting three factors for each individual sample: technical variation between samples (using spiked positive control RNA), background correction (using spiked negative control RNA) and RNA content variation (using 15 housekeeping genes). We have previously validated nCounter digital transcriptomics for simultaneous quantification of host immune and viral transcripts[36], including respiratory viruses in nasopharyngeal aspirates, even with low RNA yield[37–39].

DNA sequencing was performed on an Illumina MiSeq instrument, generating paired-end reads of 250 base pairs.

For quality control, reads were demultiplexed with LotuS v1.565[40] and processed following the DADA2 microbiome pipeline using the R packages DADA2[41] and phyloseq[42]. Briefly, reads were filtered and trimmed using the parameters truncQ=11, truncLen=c(130,200), and trimLeft=c(30, 30) and then denoised. After removing chimeras, amplicon sequence variants (ASVs) table was constructed and taxonomy was assigned using the Ribosomal Database Project (RDP) classifier implemented in DADA2 (RDP trainset 16/release 11.5). The abundance table was then corrected for copy number, rarefied to even sampling depth, and decontaminated. For decontamination, we used the prevalence-based contaminant identification method in the R package decontam[43].

16S statistical analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

All the analyses were performed using R v3.6.0 and the packages vegan[44], phyloseq[42], CoDaSeq[45], DESeq2[46], Biostrings[47], rstatix[48] and DECIPHER[49].

To analyze the 16S amplicon data, technical replicates were pooled and counts from technical replicates were added. For all the analyses using genus-level agglomerated data, only samples containing more than 10,000 reads assigned at the genus level were used (101 samples in total). Alpha-diversity was analyzed using Shannon's Diversity Index. Comparison of the alpha diversity values across different groups was performed using Wilcoxon signed-rank tests for 2-group comparisons, and Kruskal-Wallis tests for comparisons across multiple groups. In the latter case, pairwise comparisons (when applicable) were performed using Dunn post-hoc tests. To de-confound for the effect of the ICU length, we fitted a quadratic model between the days spent at ICU and the Shannon index using the lm function in R. The residuals of this model were used to test the association with the SARS-CoV-2 viral load.

Beta diversity analyses were performed using distance-based redundancy analyses (RDA), using Euclidean distances on CLR-transformed data. RDA analyses were performed using the capscale function from vegan. Non-redundant contribution to variation was calculated using the ordiR2step function from vegan, using only the variables that were significant individually in the RDA, and a null model without any explanatory variables. For these analyses, taxa with prevalence lower than 10% were excluded. Metadata variables containing dates, as well as non-informative metadata (containing a single non-NA value or unique for only one patient) were also excluded.

Differential taxa abundance analyses were performed using DESeq2's likelihood ratio tests and controlling for potential confounders when indicated, including them in the null model. All statistical tests are two-sided, and when multiple tests were applied to the different features (e.g. the differential taxa abundances across two conditions) p-values were corrected for multiple testing using Benjamini-Hochberg's method.

In order to explore species-level and strain-level diversity, 16S sequences were first clustered into 97% nucleotide diversity operational taxonomic units (OTUs) using the R

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

packages Biostrings and DECIPHER. These OTUs were used to represents the species-level. The number of unique 16S sequences clustered within each OTU were used to represent the number of detectable strains per species. To calculate strain-level diversity per sample, the number of strains of 5 detected OTU species were randomly selected and averaged. This was repeated 1,000x and the average of the all 1,000 subsamplings was used as the final strain-level diversity value for each sample, as previously described[50]. To account for uneven sampling assessing diversity differences based on different parameters, we randomly selected and averaged the species- and strain-level diversity of 5 samples per parameter. This was repeated 100x and the subsamplings were used to assess the significant differences between species- and strain-level diversity across the parameters. The average was of all 100 subsamplings was used to as the input for a Pearson's correlation between species- and strain-level diversity.

Identification of microbial reads in BAL scRNA-seq data

Single-cell data was processed with an in-house pipeline to identify microbial reads. Only read 2, containing the information on the cDNA, was used. Trimmomatic[51] (v0.38) was used to remove trim low quality bases and discard short reads. Additionally, Prinseq++[52] (v1.2) was used to discard reads with low-complexity stretches. Following quality control, reads from human and potential sequencing artifacts (phage phiX174) were mapped with STAR[53] (v2.7.1) and discarded. The remaining reads were mapped against bacterial genomes using a 2-step approach: first, we scanned the reads using mash screen[54] (v2.0) against a custom database of 11685 microbial reference genomes including bacteria, archaea, fungi and viruses. Genomes likely to be present in the analyzed sample (selected using a threshold of at least two shared hashes from mash screen) were selected and reads were pseudoaligned to this subset of genomes using kallisto[55] (v0.44.0). To remove potential artifacts, two additional filters were applied: first, if less than 10 different functions were expressed from a given species, the species was discarded. Second, if one function accounted for more than 95% of the mapped reads of a given species, it was also discarded. These filters were aimed at

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

removing potential artifacts caused by contamination or errors in the genome assemblies used in our database.

Differences in lower respiratory tract microbial taxa between COVID-19 patients and controls, ICU and ward patients, and invasive and non-invasive ventilation types were calculated using Wilcoxon rank-sum tests on centered-log-ratio (CLR)-transformed data. Prior to CLR data transformation, we filtered the data using the CoDaSeq.filter function, to keep samples with more than 1,000 reads and taxa with a relative abundance above 0.1%. Zeros were imputed using the minimum proportional abundance detected for each taxon. This more lenient approach than the one used for 16S data was chosen due to the low number of samples available and the reduced number of bacterial reads identified per sample.

Bacterial reads were assigned their specific barcodes and UMIs as follows: IDs from the mapped microbial reads were retrieved from the kallisto pseudobam output, and used to retrieve their specific barcodes and UMIs using the raw data files from read 1, assigning each barcode and UMI univocally to a microbial species and function.

Direct associations between bacteria and host cells

Host single-cell transcriptomics data was obtained from the Seurat[56] object after preprocessing and integrating the samples of the single-cell cohort, as described previously[22]. From the Seurat object, the metadata was extracted, including the information on patient group (COVID-19 or control) and severity of the disease (moderate or critical) as well as cell type and subtype annotation corresponding to each barcode. Enrichment of bacteria detected in patient groups or cell types was calculated using chi-squared tests, with effect sizes determined via the standardized residuals. Significance was assessed via post-hoc tests using the R package chisq.posthoc.test[57].

For cell types showing an enrichment in associated bacteria, a new Seurat object was created by subsetting the specific cell type. Chi-squared tests were also used to

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

determine enrichment of bacteria-associated cell subtypes. Previous annotations of cell subtypes[22] were used to generate new clusters manually and identify marker genes for these subtypes, using the function findAllMarkers from Seurat. This function was also used to find differentially expressed genes between bacteria-associated and not-bacteria-associated host cells of each subtype. When using this function, reported adjusted p-values are calculated using Bonferroni correction by default.

Figure legends

Figure 1. Sample overview and alpha diversity. a) Longitudinal sampling of patients. Each line represents one patient. Yellow lines span the days spent in ward, while blue lines span the days spent in ICU. Red points mark hospital discharge dates. Crosses indicate the timepoints were swab samples were obtained for microbiome analyses. b) Top 15 most abundant genera in this cohort. Samples with > 10,000 reads assigned to microbial taxa at the genus level were stratified according to the sampling moment: upon admission, throughout the ICU stay or at ICU discharge/during treatment in ward. c) Correlation between the SARS-CoV-2 viral load and Shannon diversity index of all samples. The shaded area surrounding the trend line represents the 95% confidence interval. d) Shannon diversity index of all samples, stratified by the sampling moment: admission, throughout ICU stay or at ICU discharge/during treatment in ward. Boxplots span from the first until the third quartile of the data distribution, and the horizontal line indicates the median value of the data. The whiskers extend from the quartiles until the last data point within 1.5 times the interquartile range, with outliers beyond. Individual data points are also represented.

Figure 2. Upper respiratory microbiome covariates in COVID-19. a) Significant covariates explaining microbiota variation in the upper respiratory tract in this cohort. Individual covariates are listed on the y-axis, their color corresponds to the metadata category they belong to: technical data, disease-related, microbiological tests, comorbidities or host cell populations or gene expression, the latter measured with nCounter (see Methods). Darker colors refer to the individual variance explained by each of these covariates assuming independency, while lighter colors represent the

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

cumulative and non-redundant variance explained by incorporating each variable to a model using a stepwise RDA analysis. The black horizontal line separates those variables that are significant in the non-redundant analysis on top (Patient ID and oxygen support) from the non-significant ones. b) RDA ordination plot showing the first 2 constrained axes. Ordination is constrained by the two significant variables "Patient ID" and "Oxygen support". Samples are depicted as points, whose color indicates the oxygen support type of the patient and whose shape indicates stay at ward or ICU (at the moment of sampling). Axes indicate the variance explained by the first two constrained components of the RDA analysis. c) Species- (left) and strain-level diversity (right) of the samples, stratified by oxygen support type. d) Pearson correlation between average species- and strain- level diversity for each of the oxygen support categories. e) Significant differences among oxygen support types. Differentially abundant taxa between invasive (red) and non-invasive (blue) ventilated sample. Only the top 10 most significant taxa are shown, as determined by their adjusted p-value. Boxplots span from the first until the third quartile of the data distribution, and the horizontal line indicates the median value of the data. The whiskers extend from the quartiles until the last data point within 1.5 times the interquartile range, with outliers beyond. Individual data points are also represented. Gray lines join samples pertaining to the same patient, taken at different time points.

Figure 3. Host single cells associated to the lower respiratory tract microbiota. a) relative proportion of cells from negative and positive COVID-19 patients with (red color) and without (blue) associated bacteria. The p-value of a chi-squared test using the count data is shown on top of the panel. b) Cell types enriched in bacteria-associated cells. Barplots represent the proportion of cell types without ("No") and with ("Yes") bacteria in COVID-19 positive and negative patients. For each patient class, we tested for enrichment of bacteria-associated cells ("Yes") across the different cell types, using the proportions of non-bacteria associated cells ("No") as background. Asterisks mark the cell types with significant enrichment of bacteria. c) Bacterial genera preferentially associating to specific cell types. The heatmaps show the standardized residuals of a chi-squared test including all bacterial genera and the three host cell types enriched in bacteria, for controls (left) and COVID-19 positive patients

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

(right). Taxa with no significant associations with any of the cell types are not shown. Asterisks denote significant positive or negative associations: enrichments are shown in red; depletions are depicted in blue. d) Host cell subtypes associated with bacteria. The heatmap shows the standardized residuals of a chi-squared test including the subtypes of neutrophils, monocytes and monocyte-derived macrophages with associated bacteria, considering cells without bacteria as background. Asterisks denote significant positive or negative associations: enrichments are shown in red; depletions are depicted in blue. e) Marker genes detected for the 5 different subtypes of neutrophils. The heatmap also shows within-group differences between bacteria-associated and bacteria-non-associated cells. f) Myeloid cell functional gene set showing the expression of canonical pro-inflammatory, anti-inflammatory and MHC genes for the two subtypes of myeloid cells significantly associated with bacteria ($CCL2^{hi}$-macrophages and $IL1B^{hi}$-monocytes). The heatmap also shows within group differences between bacteria-associated and bacteria-non-associated cells. Statistically significant differences after multiple testing correction are marked with squares. For b)-d) asterisks denote significance as follows: * = p-value ≤ 0.05; ** = p-value ≤ 0.01; *** = p-value ≤ 0.001; **** = p-value ≤ 0.0001.

Supplementary Figure Legends

Supplementary Figure 1. Alpha diversity in the upper respiratory tract. a) Correlation between number of days spent in ICU at the moment of sampling and Shannon diversity index. For samples taken after discharge to ward, the total number of days spent in ICU was used. Spearman correlation and p-value are indicated in the upper right corner of the panel. The shaded area surrounding the trend line represents the 95% confidence interval. b) Shannon diversity of the samples, grouped by patient clinical status at the moment of sampling. The p-value of a Kruskal-Wallis test is shown in the upper right corner of the panel. c) Shannon diversity of the samples, grouped by type of oxygen support supplied to the patient at the moment of sampling. The p-value of a Kruskal-Wallis test is shown in the upper right corner of the panel. d) Correlation between the days spent in ICU at the moment of sampling and SARS-CoV-2 viral load

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

of the sample. For samples taken after discharge to ward, the total number of days spent in ICU was used. The shaded area surrounding the trend line represents the 95% confidence interval. e) Correlation between SARS-CoV-2 viral load and Shannon diversity index, after controlling for the time spent in ICU. The residuals of a quadratic fit between the Shannon diversity and the days in ICU were correlated to the SARS-CoV-2 viral loads measured in the samples. Spearman correlation and p-value are indicated in the upper right corner of the panel. The shaded area surrounding the trend line represents the 95% confidence interval. For b) and c), boxplots span from the first until the third quartile of the data distribution, and the horizontal line indicates the median value of the data. The whiskers extend from the quartiles until the last data point within 1.5 times the interquartile range, with outliers beyond. Individual data points are also represented.

Supplementary Figure 2. Differentially abundant taxa between oxygen support types. The 32 taxa whose abundance is significantly different between non-invasive and invasive ventilation are represented. Boxplots span from the first until the third quartile of the data distribution, and the horizontal line indicates the median value of the data. The whiskers extend from the quartiles until the last data point within 1.5 times the interquartile range, with outliers beyond. Individual data points are also represented. Gray lines join samples pertaining to the same patient, taken at different time points.

Supplementary Figure 3. Absolute microbial read counts in single-cell RNA-seq data from BAL samples. The top 15 species detected in our analyses are depicted. Samples are grouped by disease type (control for non-COVID-19 pneumonia patients, or COVID-19) and hospital stay (ICU or ward).

Supplementary Figure 4. Associations of specific cell types with bacteria, for COVID-19 and control samples. The colors represent the strength of the association as the standardized residuals of a Chi-squared test. Red colors indicate a positive association (i.e. enrichment) of bacteria for each cell type. Blue colors indicate a negative association (i.e. depletion) of bacteria for a given cell type. Asterisks denote

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

significance as follows: * = p-value ≤ 0.05; ** = p-value ≤ 0.01; *** = p-value ≤ 0.001; **** = p-value ≤ 0.0001.

Author contributions

VLR, ACG, JW, JR designed the study. SJ, TVW, JN, CD, JG, GH, PM collected and processed the BAL samples. PVM and LV collected the clinical data. JW and EW collected the swabs. VLR, ACG and JVW processed the swabs. JVW, MB and SMM determined SARS-CoV-2 viral loads and host gene expression from swabs. DL and JQ generated the single-cell raw data as well as the processed gene-count matrix with annotations of cell types and subtypes. VLR and AGC analyzed the data. VLR, ACG and JR wrote the manuscript. All authors have read and approved the manuscript.

Acknowledgments

This study has been supported by funding from the VIB Grand Challenges Program. VLR is supported by an FWO senior postdoctoral fellowship (12V9421N). ACG is supported by an EMBO postdoctoral fellowship (ALTF 349-2019). The Raes lab is supported by KU Leuven, the Rega institute and VIB.

Conflict of interest declaration

The authors declare no competing interests.

CONTAGIOUS collaborators

Yannick Van Herck, Alexander Wilmer, Michael Casaer, Stephen Rex, Nathalie Lorent, Jona Yserbyt, Dries Testelmans, Karin Thevissen.

Data availability

Raw amplicon sequencing data that support the findings of this study have been deposited at the European Genome-phenome Archive (EGA), with accession no EGAS00001004951.

References

1. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054–1062 (2020).

2. Grasselli, G. *et al.* Risk Factors Associated with Mortality among Patients with COVID-19 in Intensive Care Units in Lombardy, Italy. *JAMA Intern. Med.* **180**, 1345–1355 (2020).
3. Mikami, T. *et al.* Risk Factors for Mortality in Patients with COVID-19 in New York City. *J. Gen. Intern. Med.* 1–10 (2020). doi:10.1007/s11606-020-05983-z
4. Guo, W. *et al.* Diabetes is a risk factor for the progression and prognosis of <scp>COVID</scp> -19. *Diabetes. Metab. Res. Rev.* **36**, (2020).
5. Lighter, J. *et al.* Obesity in Patients Younger Than 60 Years Is a Risk Factor for COVID-19 Hospital Admission. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **71**, 896–897 (2020).
6. Yu, X. *et al.* SARS-CoV-2 viral load in sputum correlates with risk of COVID-19 progression. *Crit. care* **24**, 170 (2020).
7. Magleby, R. *et al.* Impact of Severe Acute Respiratory Syndrome Coronavirus 2 Viral Load on Risk of Intubation and Mortality Among Hospitalized Patients With Coronavirus Disease 2019. *Clin. Infect. Dis.* (2020). doi:10.1093/cid/ciaa851
8. Westblade, L. F. *et al.* SARS-CoV-2 Viral Load Predicts Mortality in Patients with and without Cancer Who Are Hospitalized with COVID-19. *Cancer Cell* (2020). doi:10.1016/j.ccell.2020.09.007
9. Coperchini, F., Chiovato, L., Croce, L., Magri, F. & Rotondi, M. The cytokine storm in COVID-19: An overview of the involvement of the chemokine/chemokine-receptor system. *Cytokine and Growth Factor Reviews* **53**, 25–32 (2020).
10. Henderson, L. A. *et al.* On the Alert for Cytokine Storm: Immunopathology in COVID-19. *Arthritis and Rheumatology* **72**, 1059–1063 (2020).
11. Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and immunosuppression. *The Lancet* **395**, 1033–1034 (2020).
12. Khatiwada, S. & Subedi, A. Lung microbiome and coronavirus disease 2019 (COVID-19): Possible link and implications. *Human Microbiome Journal* **17**, 100073 (2020).
13. Dickson, R. P., Martinez, F. J. & Huffnagle, G. B. The role of the microbiome in exacerbations of chronic lung diseases. *The Lancet* **384**, 691–702 (2014).
14. Huffnagle, G. B., Dickson, R. P. & Lukacs, N. W. The respiratory tract microbiome and lung inflammation: A two-way street. *Mucosal Immunology* **10**, 299–306 (2017).
15. Zhang, H. *et al.* Metatranscriptomic Characterization of COVID-19 Identified A Host Transcriptional Classifier Associated With Immune Signaling. *Clin. Infect. Dis.* (2020). doi:10.1093/cid/ciaa663
16. Xu, R. *et al.* Temporal dynamics of human respiratory and gut microbiomes during the course of COVID. *medRxiv* 2020.07.21.20158758 (2020). doi:10.1101/2020.07.21.20158758
17. Han, Y., Jia, Z., Shi, J., Wang, W. & He, K. The active lung microbiota landscape of COVID-19 patients. *medRxiv* 2020.08.20.20144014 (2020). doi:10.1101/2020.08.20.20144014
18. Mostafa, H. H. *et al.* Metagenomic Next-Generation Sequencing of Nasopharyngeal Specimens Collected from Confirmed and Suspect COVID-19

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

Patients Downloaded from. *MBio* **11**, (2020).

19. Ho Man, W., de Steenhuijsen Piters, W. A. & Bogaert, D. The microbiota of the respiratory tract: gatekeeper to respiratory health. (2017). doi:10.1038/nrmicro.2017.14
20. Whelan, F. J. *et al.* Longitudinal sampling of the lung microbiota in individuals with cystic fibrosis. *PLoS One* **12**, (2017).
21. Nolan, T. J. *et al.* Low-pathogenicity Mycoplasma spp. alter human monocyte and macrophage function and are highly prevalent among patients with ventilator-acquired pneumonia. *Thorax* **71**, 594–600 (2016).
22. Wauters, E. *et al.* Discriminating Mild from Critical COVID-19 by Innate and Adaptive Immune Single-cell Profiling of Bronchoalveolar Lavages. *Patrick Matthys* **9**, 14 (2020).
23. Harris, J. R., Balajee, S. A. & Park, B. J. Pneumocystis jirovecii pneumonia: Current knowledge and outstanding public health issues. *Current Fungal Infection Reports* **4**, 229–237 (2010).
24. Wang, S. *et al.* S100A8/A9 in inflammation. *Frontiers in Immunology* **9**, 1298 (2018).
25. Coveney, A. P. *et al.* Myeloid-related protein 8 induces self-tolerance and cross-tolerance to bacterial infection via TLR4- and TLR2-mediated signal pathways. *Nat. Publ. Gr.* (2015). doi:10.1038/srep13694
26. Holm, J. & Hansen, S. I. Characterization of soluble folate receptors (folate binding proteins) in humans. Biological roles and clinical potentials in infection and malignancy. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1868**, 140466 (2020).
27. Zakharkina, T. *et al.* The dynamics of the pulmonary microbiome during mechanical ventilation in the intensive care unit and the association with occurrence of pneumonia. doi:10.1136/thoraxjnl-2016-209158
28. Schmitt, F. C. F. *et al.* Pulmonary microbiome patterns correlate with the course of disease in patients with sepsis-induced ARDS following major abdominal surgery. (2020). doi:10.1016/j.jhin.2020.04.028
29. Shen, Z. *et al.* Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients With Coronavirus Disease 2019. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **71**, 713–720 (2020).
30. Marotz, C. *et al.* Title: Microbial context predicts SARS-CoV-2 prevalence in patients and the hospital built environment. *medRxiv* 2020.11.19.20234229 (2020). doi:10.1101/2020.11.19.20234229
31. Bassis, C. M. *et al.* Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *MBio* **6**, (2015).
32. Brennan, M. T. *et al.* The role of oral microbial colonization in ventilator-associated pneumonia. *Oral Surgery, Oral Med. Oral Pathol. Oral Radiol. Endodontology* **98**, 665–672 (2004).
33. Stonecypher, K. Ventilator-Associated Pneumonia: The Importance of Oral Care in Intubated Adults. *Crit. Care Nurs. Q.* **33**, 339–347 (2010).
34. Segal, L. N. *et al.* Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat. Microbiol.* **1**, 1–11 (2016).
35. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).

36. Moens, B. *et al.* Simultaneous RNA quantification of human and retroviral genomes reveals intact interferon signaling in HTLV-1-infected CD4+ T cell lines. *Virol. J.* **9**, 171 (2012).
37. Fukutani, K. F. *et al.* Pathogen transcriptional profile in nasopharyngeal aspirates of children with acute respiratory tract infection. *J. Clin. Virol.* **69**, 190–196 (2015).
38. Bouzas, M. L. *et al.* Diagnostic accuracy of digital RNA quantification versus real-time PCR for the detection of respiratory syncytial virus in nasopharyngeal aspirates from children with acute respiratory infection. *J. Clin. Virol.* **106**, 34–40 (2018).
39. Fukutani, K. F. *et al.* In situ immune signatures and microbial load at the nasopharyngeal interface in children with acute respiratory infection. *Front. Microbiol.* **9**, (2018).
40. Hildebrand, F., Tadeo, R., Voigt, A. Y., Bork, P. & Raes, J. LotuS: An efficient and user-friendly OTU processing pipeline. *Microbiome* **2**, 30 (2014).
41. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
42. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, e61217 (2013).
43. Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
44. Oksanen, J. *et al.* vegan: Community Ecology Package. (2019).
45. Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V. & Egozcue, J. J. It's all relative: analyzing microbiome data as compositions. *Ann. Epidemiol.* **26**, 322–329 (2016).
46. 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).
47. Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings: Efficient manipulation of biological strings. (2019).
48. Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. (2020).
49. Wright, E. S. *Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R*.
50. Gregory, A. C. *et al.* Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**, 1109-1123.e14 (2019).
51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
52. Cantu, V. A., Sadural, J. & Edwards, R. PRINSEQ++, a multi-threaded tool for fast 1 and efficient quality control and 2 preprocessing of sequencing datasets. (2019). doi:10.7287/peerj.preprints.27553v1
53. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
54. Ondov, B. D. *et al.* Mash Screen: high-throughput sequence containment

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

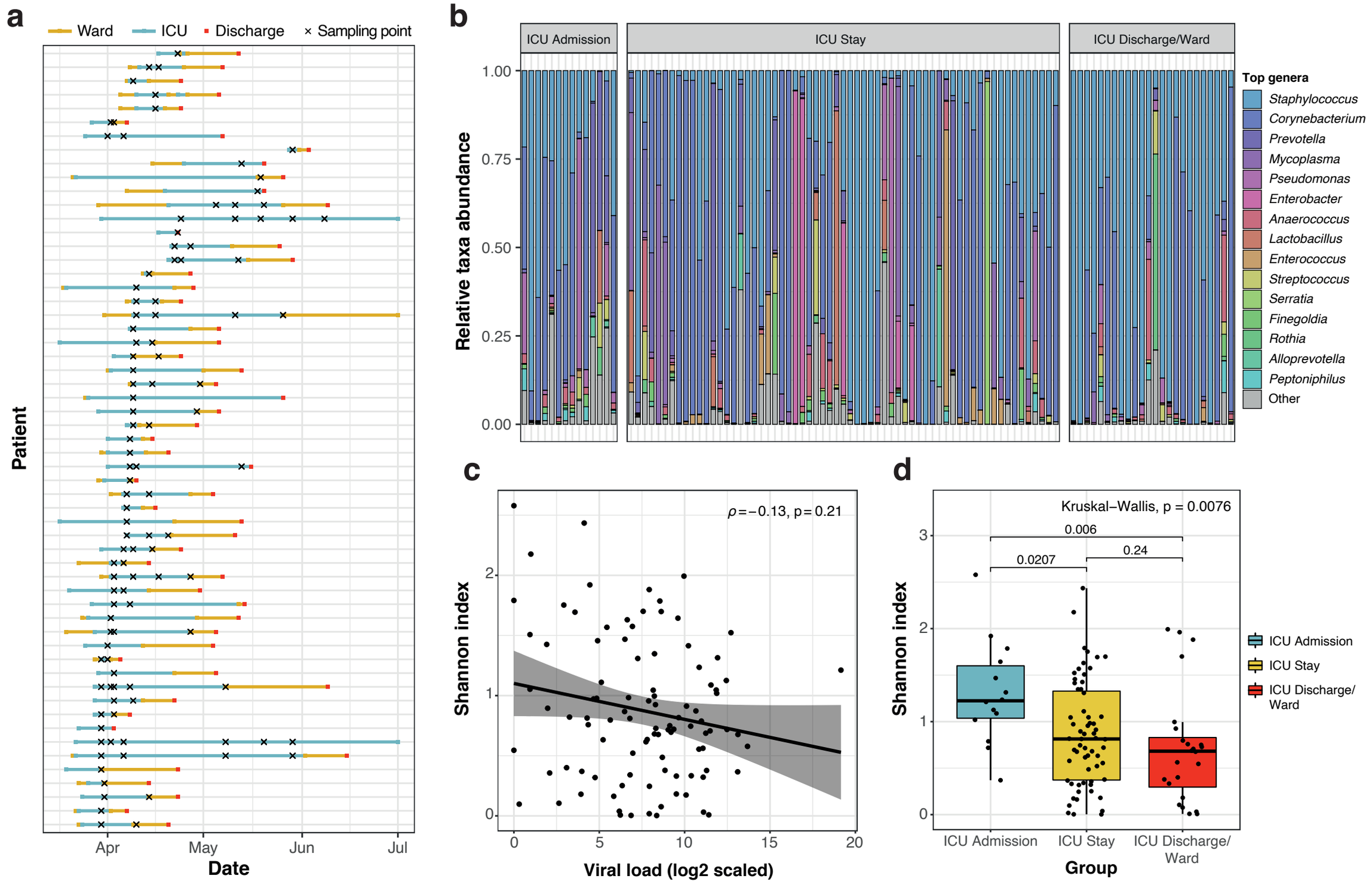
estimation for genome discovery. *Genome Biol.* **20**, 232 (2019).

55. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

56. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

57. Ebbert, D. chisq.posthoc.test: A Post Hoc Analysis for Pearson's Chi-Squared Test for Count Data. (2019).

Figure 1


a
Ward
ICU
Discharge
× Sampling point
Patient
Apr
May
Jun
Jul
Date
b
ICU Admission
ICU Stay
ICU Discharge/Ward
Relative taxa abundance
1.00
0.75
0.50
0.25
0.00
Top genera
Staphylococcus
Corynebacterium
Prevotella
Mycoplasma
Pseudomonas
Enterobacter
Anaerococcus
Lactobacillus
Enterococcus
Streptococcus
Serratia
Finegoldia
Rothia
Alloprevotella
Peptoniphilus
Other
c
ρ = −0.13, p = 0.21
Shannon index
Viral load (log2 scaled)
d
Kruskal–Wallis, p = 0.0076
0.006
0.0207
0.24
Shannon index
ICU Admission
ICU Stay
ICU Discharge/ Ward
Group

It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2

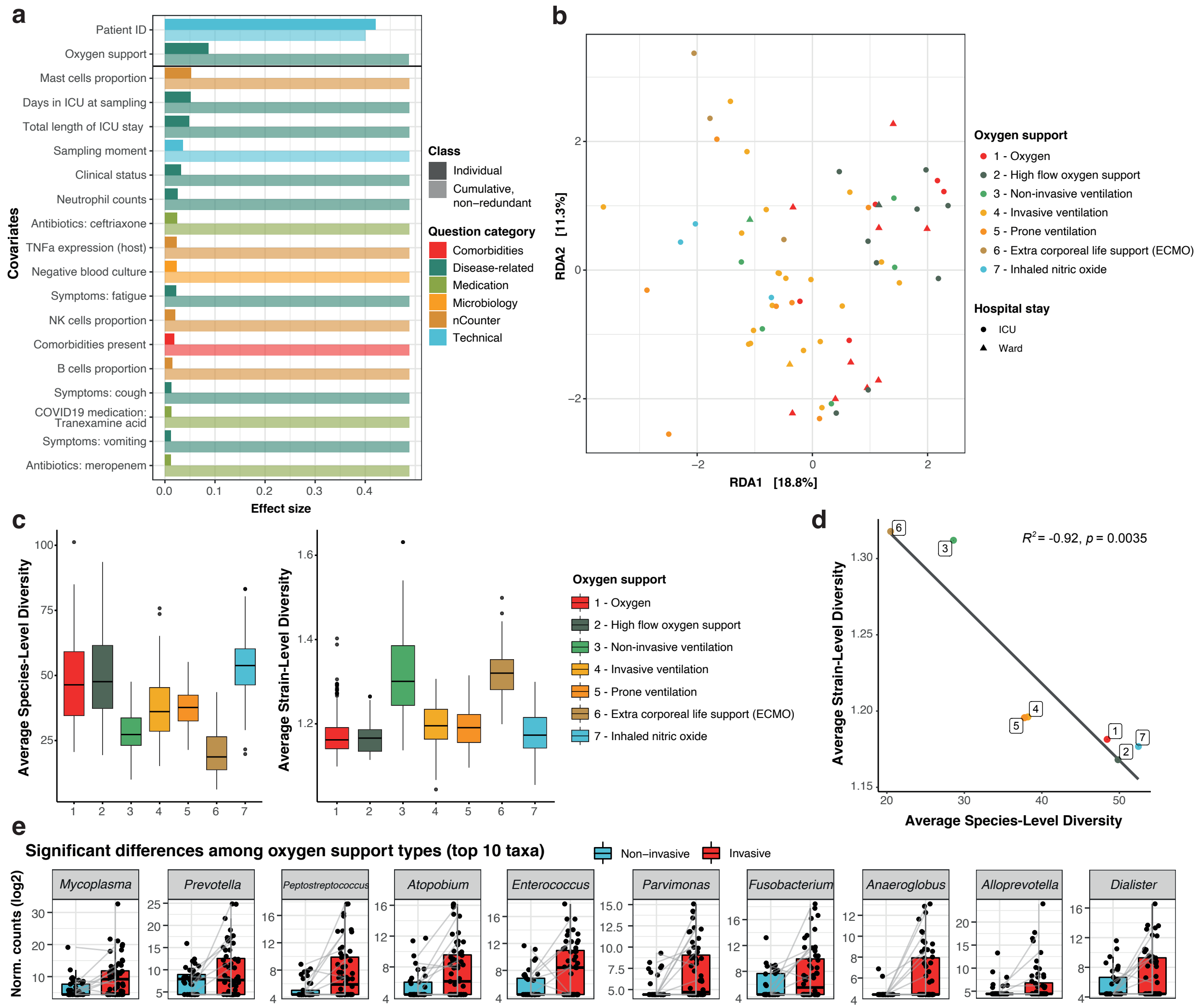
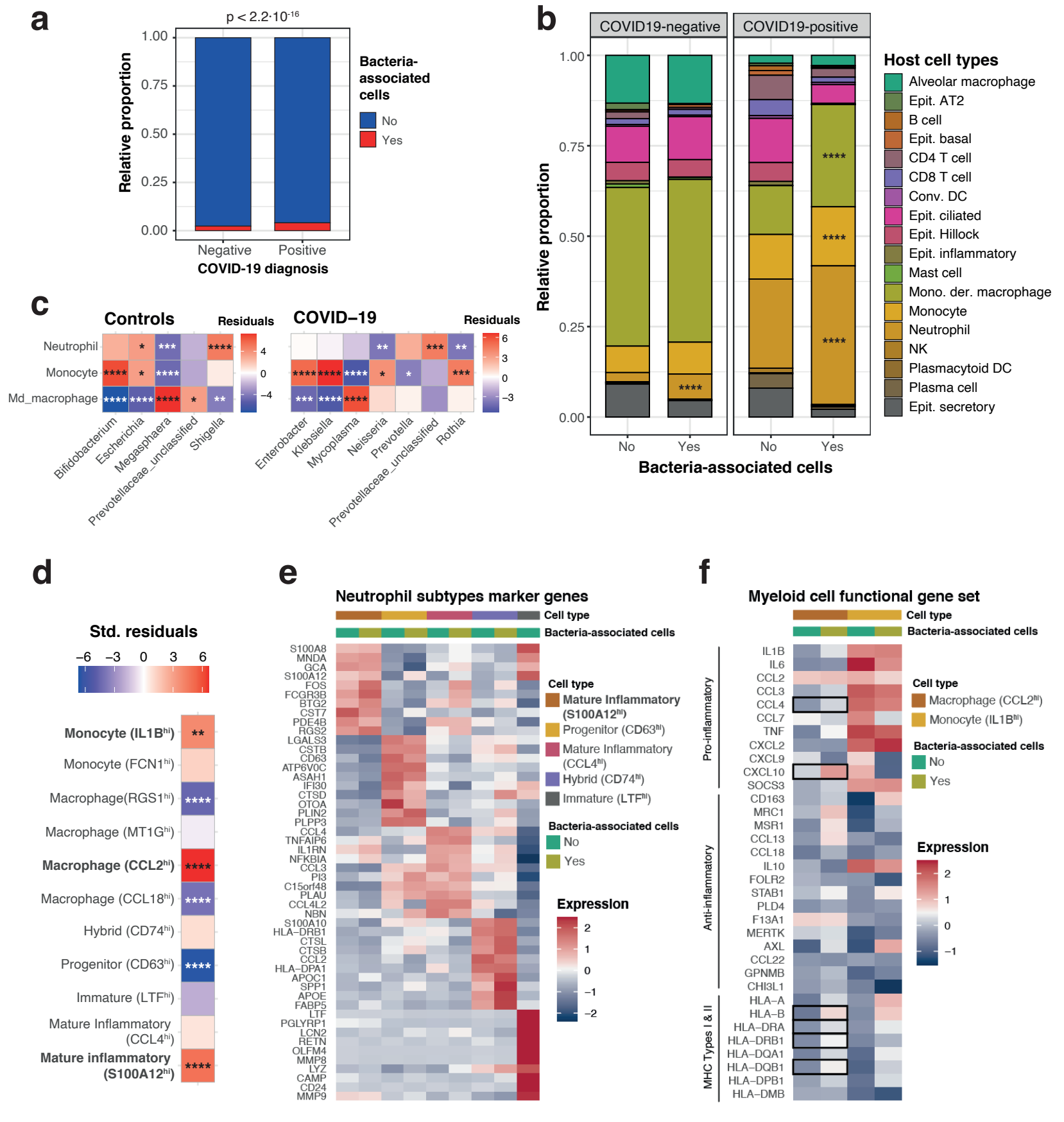


Figure 3


a
p < 2.2·10-16
Relative proportion
1.00
0.75
0.50
0.25
0.00
Negative
Positive
COVID-19 diagnosis
Bacteria-associated cells
No
Yes
b
COVID19-negative
COVID19-positive
Relative proportion
Bacteria-associated cells
Host cell types
Alveolar macrophage
Epit. AT2
B cell
Epit. basal
CD4 T cell
CD8 T cell
Conv. DC
Epit. ciliated
Epit. Hillock
Epit. inflammatory
Mast cell
Mono. der. macrophage
Monocyte
Neutrophil
NK
Plasmacytoid DC
Plasma cell
Epit. secretory
c
Controls
COVID–19
Residuals
Neutrophil
Monocyte
Md_macrophage
Bifidobacterium
Escherichia
Megasphaera
Prevotellaceae_unclassified
Shigella
Enterobacter
Klebsiella
Mycoplasma
Neisseria
Prevotella
Prevotellaceae_unclassified
Rothia
d
Std. residuals
Monocyte (IL1Bhi)
Monocyte (FCN1hi)
Macrophage(RGS1hi)
Macrophage (MT1Ghi)
Macrophage (CCL2hi)
Macrophage (CCL18hi)
Hybrid (CD74hi)
Progenitor (CD63hi)
Immature (LTFhi)
Mature Inflammatory (CCL4hi)
Mature inflammatory (S100A12hi)
e
Neutrophil subtypes marker genes
Cell type
Bacteria-associated cells
Mature Inflammatory (S100A12hi)
Progenitor (CD63hi)
Mature Inflammatory (CCL4hi)
Hybrid (CD74hi)
Immature (LTFhi)
No
Yes
Expression
f
Myeloid cell functional gene set
Pro-inflammatory
Anti-inflammatory
MHC Types I & II
Macrophage (CCL2hi)
Monocyte (IL1Bhi)
Expression

medRxiv preprint doi: https://doi.org/10.1101/2020.12.23.20248425; this version posted February 7, 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.
It is made available under a CC-BY-NC-ND 4.0 International license .

	Upper respiratory tract (swabs)	***Lower respiratory tract (BAL)***
Number of patients	58	35
COVID-19 positive (%)	58 (100%)	22 (62.9%)
Age (range)	61.2 (37-83)	64.1 (45-85)
Female sex (%)	13 (22.4%)	12 (34.3%)
BMI (range)	28.9 (22-46.7)	26.2 (16-36.4)
Diabetic (%)	12 (20.7%)	6 (17.1%)
Days in ICU (range)	21.4 (2-72)	NA
Days in hospital (range)	32.5 (6-86)	NA

Table 1. Patient demographics of our upper and lower respiratory tract cohorts