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

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

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.

95 To identify potential associations between COVID-19 severity and evolution and the
96 upper and lower respiratory tract microbiota, we used nasopharyngeal swabs and
97 bronchoalveolar lavage (BAL) samples, respectively. For the upper respiratory tract, we
98 longitudinally profiled the nasopharyngeal microbiome of 58 COVID-19 patients during
99 intensive care unit (ICU) treatment and after discharge to a classical hospital ward
100 following clinical improvement, in conjunction with viral load determination and
101 nCounter immune profiling. For the lower respiratory tract, we analyzed microbial
102 signals in cross-sectional single-cell RNA-seq data from of bronchoalveolar lavage (BAL)
103 samples of 22 COVID-19 patients and 13 pneumonitis controls with negative COVID-19
104 qRT-PCR, obtained from the same hospital. The integration of these data enabled us to
105 (1) identify potential confounders of COVID-19 microbiome associations, (2) explore
106 how microbial diversity evolves throughout hospitalization, (3) study microbe-host cell
107 interaction and (4) substantiate a link between the respiratory microbiome and SARS-
108 CoV-2 viral load as well as COVID-19 disease severity. Altogether, our results directly
109 point to specific interactions between the microbiota and the immune cells, likely
110 driven by clinical ventilation practices, which could potentially influence COVID-19
111 disease progression and resolution.
112
113 **Results**
114
115 **The upper respiratory microbiota of COVID-19 patients**
116
117 We longitudinally profiled the upper respiratory microbiota of 58 patients diagnosed
118 with COVID-19 based on a positive qRT-PCR test or a negative test with high clinical
119 suspicion based on symptomatology and a chest CT-scan showing alveolar damage. All
120 these patients were admitted and treated at UZ Leuven hospital. Patient demographics
121 for this cohort are shown in Table 1.
122
123 In total, 112 nasopharyngeal swabs from these patients were processed (Figure 1a):
124 the V4 region of the 16S rRNA gene amplified on extracted DNA using 515F and 806R
125 primers, and sequenced on an Illumina MiSeq platform (see Methods). From the same
126 swabs, RNA was extracted to determine SARS-CoV-2 viral loads and to estimate

127 immune cell populations of the host using nCounter (Methods). Of the 112 samples
128 processed and sequenced, 101 yielded over 10,000 amplicon reads that could be
129 assigned to bacteria at the genus level (Figure 1b; Methods). The microbiome of the
130 entire cohort was dominated by the gram-positive genera *Staphylococcus* and
131 *Corynebacterium*, typical from the nasal cavity and nasopharynx[19].
132
133 **Bacterial alpha diversity is strongly associated with ICU stay length**
134
135 First, we determined genus-level alpha-diversity for the 101 samples with more than
136 10,000 assigned reads, using Shannon Diversity index (see Methods; Supplementary
137 Table 1). We observed that alpha diversity was not significantly correlated to SARS-
138 CoV-2 viral load in nasopharyngeal swabs (Figure 1c). In contrast, we found the
139 Shannon index to be significantly affected by the sampling moment (Kruskal-Wallis
140 test, p-value = 0.0076; Figure 1d), with significant differences between swabs procured
141 upon patient admission and later timepoints, suggesting an important effect of disease
142 progression and/or treatment (i.e. due to antibiotics administered throughout ICU
143 stay). We explored these differences further, and observed that Shannon Diversity
144 index correlated negatively with the number of days spent in ICU at the moment of
145 sampling, with longer ICU stays leading to a lower diversity (Supplementary Figure 1a;
146 ρ=-0.55, p-value=$4.4 \cdot 10^{-9}$). Furthermore, the observed decrease in diversity occurred
147 mostly over the first 2 weeks in ICU (Supplementary Figure 1a). Other severity
148 indicators, such as the patient clinical status (i.e. a qualitative metric used to classify
149 patients into different levels of disease severity) or the type of oxygen support
150 required at the moment of sampling, showed no association with the genus-level
151 Shannon index (Supplementary Figure 1b,c). Therefore, the differences in diversity
152 observed across samples were likely driven by the time spent in ICU, and not
153 specifically by disease progression as other severity indicators were unaffected.
154
155 Since time in ICU had a major effect in alpha diversity, we explored whether it might
156 be masking any effects of viral load, as SARS-CoV-2 load was not associated with the
157 length of the ICU stay (Supplementary Figure 1d). When controlling for the time in ICU
158 (see Methods), we observed that viral load was negatively associated with the

159 Shannon diversity index (ρ=-0.26, p-value=0.0089; Supplementary Figure 1e). Other
160 severity indicators such as the clinical status and the type of oxygen support were not
161 correlated to diversity even after controlling for the confounding effect of ICU stay
162 (Kruskal-Wallis test, p-value>0.05).
163
164 **Microbiome composition variation is driven by the type of respiratory support**
165
166 We next explored potential associations between the upper respiratory genus-level
167 microbiota composition and the extensive metadata collected in the study. In total, 70
168 covariates related to patient anthropometrics, medication and clinical variables
169 measured in the hospital and host cytokine expression measured in the swabs were
170 tested (Supplementary Table 2). Individually, 19 of these covariates showed a
171 significant correlation to microbiota composition (dbRDA, p-value<0.05;FDR<0.05;
172 Figure 2a). These significant covariates were related to disease and measures of its
173 severity, such as the clinical evaluation of the patient, the total length of the ICU stay,
174 the number of days in ICU at the time of sampling, or the type of oxygen support
175 needed by the patient. Surprisingly, the total SARS-CoV-2 viral load detected in the
176 swabs was not significantly associated to microbiome composition variation
177 (Supplementary Table 2).
178
179 Of the 19 significant covariates, only 2 accounted for 48.7% non-redundant variation in
180 this dataset, with the rest holding redundant information. These were the patient ID,
181 included due to the longitudinal sampling of patients, and confirming that intra-
182 individual variation over time is smaller than patient inter-individual variation[20], and
183 the type of oxygen support received at the time of sampling (Figure 2a,b). Notably, the
184 type of oxygen support discriminated samples based on ventilation type, with non-
185 invasive ventilation samples (groups 1, 2 and 3) separating from samples from
186 intubated patients (groups 4 to 7; PERMANOVA test, R^2=0.0642, p-value=0.001).
187
188 To determine if oxygen support also impacted the microbiome at finer taxonomic
189 resolution, we revisited alpha-diversity at species- and strain-level. We defined species
190 as 97% identity 16S OTUs and strains per species as the clustered 16S sequences

191 within each OTU. The species and strain-level diversity per sample were calculated as
192 the number of OTUs and as the mean of the number of strains from five randomly
193 sampled OTU species sampled 1,000 times, respectively. Our analyses revealed both
194 species- and strain-level diversity change with ventilation, even with non-invasive
195 ventilation (e.g. BIPAP, CPAP). Across all samples we observed high species- and low
196 strain-level diversity pre-ventilation, which reversed following any form of ventilation
197 (Figure 2c; Wilcoxon test; p-values<0.05, with the exception of type 7), with the
198 exception of ventilation with inhaled nitric oxide. Further, Species- and strain-level
199 diversity showed a strong inverse correlation (Figure 2d; Pearson's correlation, R^2 = -
200 0.92, p-value = 0.0035).
201
202 Therefore, we evaluated which specific taxa were differentially abundant between
203 samples from intubated and non-intubated patients. In total, 30 genera were more
204 abundant in intubated samples, while 2 genera were more abundant in non-invasively
205 ventilated patients (p-value<0.05;FDR<0.05; Figure 2e, Supplementary Figure 2;
206 Supplementary Table 3). Some of these taxa are common oral microbiome
207 commensals or opportunistic pathogens that had been repeatedly reported as more
208 abundant in COVID-19 patients than in healthy controls, such as *Prevotella*, *Veillonella*,
209 *Fusobacterium*, *Porphyromonas* or *Lactobacillus*[15–17]. Here, we reported higher
210 abundance of these genera in intubated COVID-19 patients as compared to non-
211 mechanically ventilated patients. This points at mechanical ventilation as a potential
212 confounder of previous COVID-19 studies. Additionally, we found other taxa not
213 previously reported in previous COVID-19 microbiome studies, such as *Mycoplasma* or
214 *Megasphaera* (Figure 3c, Supplementary Figure 2), but previously associated to risk of
215 ventilator-associated pneumonia[21].
216
217 By extracting the amplicon sequence variants (ASVs) corresponding to these
218 differentially abundant genera (see Methods), some of these taxa could be narrowed
219 down to the species level, confirming their origin as typically oral bacteria: for
220 instance, *Prevotella* species included *P. oris*, *P. salivae*, *P. denticola*, *P. buccalis* and *P.*
221 *oralis*. Within the *Mycoplasma* genus, ASVs were assigned to *Mycoplasma salivarium*
222 among other species, an oral bacterium which has been previously associated to the

223 incidence of ventilator-associated pneumonia[21]. When controlling for ventilation type,
224 no taxa were found associated to SARS-CoV-2 viral loads. These results show that
225 further research with larger cohorts and controlling for the relevant confounders
226 highlighted here, such as ventilation type or length of stay in ICU, will be needed to
227 study the specific effect of the viral infection.
228
229 **Single-cell RNA-seq identifies oral commensals and opportunist pathogens in the**
230 **lower respiratory tract**
231
232 Next, we explored what the functional consequences of (ventilation-driven) lung
233 microbiome disturbances could be. To do so, we screened single-cell RNA-seq data
234 generated on BAL samples of 35 patients[22] to identify microbial reads. All patients in
235 this cross-sectional cohort had clinical symptoms of pneumonia, 22 of them being
236 diagnosed with COVID-19. The other 13 patients with non-COVID-19 pneumonia were
237 hereafter referred to as controls (Table 1). Microbiome read screening of these
238 samples revealed an average of 7,295.3 microbial reads per sample (ranging from 0 to
239 74,226 reads, with only a single sample yielding zero microbial reads; Supplementary
240 Figure 3).
241
242 Among the top taxa encountered in these patients, we found some similarities with
243 the data obtained in nasopharyngeal swabs. The top 15 species detected include
244 *Mycoplasma salivarium* as the dominating taxon in 5 COVID-19 patients in ICU, as well
245 as different *Prevotella* members. Non-COVID-19 pneumonia patients in ward (i.e. non-
246 intubated) harbored different microbes: 2 patients had a microbiome dominated by
247 *Porphyromonas gingivalis*, while a single patient had a microbiome dominated by the
248 fungus *Pneumocystis jirovecii*, a known pathogen causing pneumonia[23].
249 Supplementary table 4 shows associations between organism abundances and specific
250 patient metadata: disease, hospital stay and ventilation type. Multiple links with
251 COVID-19 diagnosis were identified (Wilcoxon test, (noncorrected) p-value<0.05; see
252 Methods) but due to the low sample number, none was significant after multiple-test
253 correction. Additionally, as hospital stay (ICU or ward), type of oxygen support
254 (invasive or non-invasive ventilation) and disease (COVID-19 or controls) were highly

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.

286 We also explored whether host-associated bacterial reads would preferentially be
287 linked with specific cell types, taking into account the varying frequencies of cell types
288 in COVID-19 patients and controls (see Methods). Such a preferential association
289 would suggest that these observations are biologically relevant and not an artifact of
290 the single-cell sample and library preparation. Among control patients, cell types were
291 similarly distributed in both groups (i.e. with and without bacteria), with only a
292 preferential association of microbial cells with neutrophils (p-value = $3.61 \cdot 10^{-12}$; Figure
293 3b; Supplementary Figure 4). However, in COVID-19 patients, three cell types were
294 significantly associated with bacteria: neutrophils (p-value < $2.2 \cdot 10^{-16}$), monocytes ((p-
295 value = $4.82 \cdot 10^{-5}$) and monocyte-derived macrophages (p-value < $2.2 \cdot 10^{-16}$; Figure 3b;
296 Supplementary Figure 4). We also found that different bacteria associate with distinct
297 host cells. For instance, in COVID-19 patients, bacteria from the *Mycoplasma* genus
298 preferentially associated to monocyte-derived macrophages (p-value = $2.28 \cdot 10^{-7}$),
299 while *Rothia* (p-value = $8.21 \cdot 10^{-4}$), *Enterobacter* (p-value = $2.59 \cdot 10^{-5}$), or *Klebsiella* (p-
300 value = $3.12 \cdot 10^{-9}$) are enriched in monocytes (Figure 3c).
301
302 Last, we investigated whether the associations of bacteria to host cells are linked to
303 host cell expression. To do so, we assessed whether expression based cell subtype
304 classification[22] for neutrophils, monocytes and macrophages showed non-random
305 associations with bacteria across all samples in this cohort. Among the neutrophils, a
306 subtype of inflammatory neutrophils characterized by expression of the calgranulin
307 S100A12 was enriched in bacteria-associated cells (p-value $7.18 \cdot 10^{-6}$; Figure 3d,e). This
308 subset of cells was also found to be enriched in SARS-CoV-2 nucleocapsid gene reads[22],
309 suggesting that the same cell type responsible for defense against the virus would be
310 responding to potentially invasive bacteria in the lung. This subgroup is characterized
311 by the expression of the calprotectin subunits S100A8 and S100A9. It is known that
312 S100A8/A9 heterodimer secretion is increased in infection-induced inflammation and
313 has some antibacterial effects mediated by secretion of pro-inflammatory cytokines,
314 release of reactive oxygen species and recruitment of other inflammatory cells, as well
315 as chelation of Zn^{2+} necessary for bacterial enzymatic activity[24]. These mechanisms are
316 mediated by binding of the S100A8/A9 dimer to TLR4 receptors to trigger the release
317 of pro-inflammatory cytokines such as IL-6 and TNF-α, and thus may contribute to

318 sustain or exacerbate inflammation[25]. Therefore, the association with bacteria could,
319 at least in part, explain the inflammatory phenotype of this neutrophil subset. To
320 further examine this hypothesis, we explored differential gene expression between
321 bacteria-associated and non-associated S100A12^{hi} neutrophils (Supplementary Table
322 6). Because association of these cells with SARS-CoV-2 and with bacteria was mutually
323 exclusive, we also compared these changes with the ones triggered by the virus in
324 neutrophils[22]. Within this subset, neutrophils with co-occurring bacteria showed
325 significantly higher expression (Bonferroni-corrected p-value < 0.05) of pro-
326 inflammatory genes, including the cytokine IL1B and some of its target genes (PTSG2),
327 the transcription factors FOS and JUN, and several genes involved in degranulation
328 (S100A9, FOLR3, HSPA1A, HSP90AA1, FCGR3B), (Supplementary Table 6). Among
329 these, FOLR3, a gene encoding for a folate receptor, is found in neutrophil secretory
330 granules and has antibacterial functions, by binding folates and thus depriving bacteria
331 of these essential metabolites[26]. This response differed to that of virus-engulfing
332 neutrophils in that IFN response genes are not distinctively upregulated by bacteria.
333
334 Regarding myeloid cells, both inflammatory IL1B^{hi} monocytes (p-value = $2 \cdot 10^{-16}$) as well
335 as a mixed group of CCL2-expressing macrophages (p-value = $5.38 \cdot 10^{-10}$) are enriched
336 in bacteria-associated cells (Figure 3f). These inflammatory monocytes are believed to
337 have an important role in the cytokine storm occurring in severe COVID-19 patients. In
338 this case, further gene expression patterns were detected, specific for bacteria-
339 associated cells: for CCL2^{hi} macrophages, cells with co-occurring bacteria expressed
340 higher levels of MHC genes of type I and II, suggesting a more active role of these cells
341 in antigen presentation (Bonferroni-corrected p-value < 0.05; Figure 3f; Supplementary
342 Table 6). A similar increase was also observed in monocytes, yet not significant
343 (Supplementary Table 6), possibly due to the lower monocyte abundances in this
344 dataset. Additionally, bacteria-associated macrophages express significantly higher
345 levels of the calprotectin subunits S100A8/A9, similarly to neutrophils, as well as pro-
346 inflammatory chemokines (such as CCL4, CXCL10 and CXCL1).
347

348 Altogether, our results suggest that the bacteria detected in these cell subsets via
349 scRNA-seq analyses may be contributing to the inflammatory response observed in the
350 host.
351
352 **Discussion**
353
354 Since the beginning of the COVID-19 pandemic, a massive global effort by the scientific
355 community was undertaken to understand physiopathology of SARS-CoV-2 infection
356 and risk factors affecting disease outcome. In this study, we explored the respiratory
357 microbiota as a potential risk factor for disease severity, and we evaluated the upper
358 and lower respiratory tract microbiota in COVID-19 patients throughout their
359 hospitalization. We linked this data to viral load measurements and immunoprofiling
360 results from nCounter and single-cell RNA sequencing data. To assess robustness of
361 previously reported signals, we investigated the effect of potential confounders based
362 on a broad panel of patient metadata variables.
363
364 We found that in the upper respiratory tract, while SARS-CoV-2 viral load has a weak
365 negative association with bacterial biodiversity, a strong effect of severity indicators
366 such as ICU stay was observed, with diversity decreasing throughout the length of the
367 ICU period, a pattern reminiscent of that seen in other pulmonary conditions[27,28].
368 This effect of ICU and/or ventilation on microbiome alpha diversity could potentially
369 explain why previous studies on the microbiota of COVID-19 patients show conflicting
370 results regarding diversity: some studies reported lower diversity in sputum or throat
371 swab samples of COVID-19 patients[15,16,18] while others focusing on the lower
372 respiratory microbiome using bronchoalveolar fluid samples, showed higher bacterial
373 diversity in COVID-19 patients than in controls[17]. To further complicate matters, it
374 cannot be excluded that sampling site or processing could also be potential
375 confounders in these studies and/or reflect the different pathologies in the different
376 areas of the respiratory tract.
377
378 We further found that between patient microbiome variation (as measured by genus-
379 level microbial beta-diversity) was also affected by different severity indicators such as

380 the clinical status of the patient, or more importantly the type of oxygen support
381 received, with intubated patients harboring a different microbiota than non-intubated
382 patients. The impact of oxygen support was also reflected at the species- and strain-
383 levels, with intubation causing a significant decrease and increase, respectively, in
384 diversity. We hypothesize that the introduction of forced oxygen may drive the fast
385 extinction of certain microbial species enabling the diversification of existing or newly
386 colonizing species into new strains. Our results suggest that non-invasive ventilation
387 (e.g. BIPAP, CPAP) can have microbial effects indicating that any form of ventilation
388 may be a tipping point for microbial community differences.
389
390 Importantly, several of the taxa reported to change between intubated and non-
391 intubated patients were reported to be linked to diagnosis in previous COVID-19
392 microbiome studies[15–17]. In our study, no taxa were differentially abundant between
393 COVID-19 positive and negative patients after controlling for patient intubation. This
394 result strongly points at the possibility that intubation and mechanical ventilation are
395 confounding previous results. Indeed, one study comparing COVID-19 patients with
396 patients diagnosed of community-acquired pneumonia found no differences in
397 respiratory microbiome composition between both groups of patients, but both
398 groups did differ from healthy controls[29]. Together, these results indicate that patient
399 intubation or even non-invasive ventilation are to be considered as important
400 confounders when studying the upper respiratory microbiome, and we strongly
401 suggest future COVID-19 microbiome studies should foresee and include strategies to
402 account for this covariate. A recent study found a single ASV corresponding to the
403 genus *Rothia* that was specific for SARS-CoV-2 patients after controlling for ICU-related
404 confounders[30].
405
406 To better understand the potential functional consequences of these procedures and
407 linked microbial shifts, we also profiled the microbiome of the lower respiratory tract
408 using single-cell data obtained from a cross-sectional cohort of patients derived from
409 the same hospital. Our results show that 'standard' single-cell RNA-seq, even though
410 not optimized for microbial detection and profiling, can identify bacteria alone or in
411 association with specific human cells. Unfortunately, the low numbers of microbial

412 reads obtained in this cohort, together with the fact that ICU stay, COVID-19 diagnosis
413 and intubation are highly correlated in this set of patients, only allow for a descriptive
414 analysis of the results. In this cohort, we identified different oral commensals and
415 opportunistic pathogens previously linked to COVID-19 patients in both groups of
416 samples, thus pointing again at a potential ventilation-linked origin. More interestingly,
417 we identified a subset of bacteria associated with host cells, more specifically with
418 neutrophils, monocytes and macrophages. This enrichment shows that these bacteria
419 are likely not random contaminants, from which an even distribution across cell types
420 (i.e. considering cell type abundances) would be expected. The identity of these host
421 cells suggests that bacteria could have been phagocyted by these innate immune
422 system cells, rather than be attached to the host cell surface. To the best of our
423 knowledge, this is the first study linking interacting host cells and lung microbiome via
424 high-throughput single-cell RNA-seq.
425
426 We find that host cells associated with bacteria, most of which are of oral origin,
427 exhibit pro-inflammatory phenotypes as well as higher levels of MHC for antigen
428 presentation. In this single-cell cohort it was observed that critical COVID-19 patients
429 are characterized by an impaired monocyte to macrophage differentiation, resulting in
430 an excess of pro-inflammatory monocytes, as well as by prolonged neutrophil
431 inflammation[22]. Given that only these cell types are enriched in bacteria, we
432 hypothesize that the respiratory (or ventilation-linked) microbiome may be playing a
433 role in exacerbating COVID-19 progression in the lower respiratory tract. We verified
434 that this response would likely be driven by bacteria and not SARS-CoV-2, which is also
435 detected mostly in these cell types, as there is almost no overlap in detection of both
436 virus and bacteria in the same cells. However, it must be noted that lack of detection
437 does not completely rule out presence of virus and bacteria within these cells.
438 Therefore, further research is required in order to confirm a causative role of the
439 microbiota in this immune impairment characteristic of critical disease, and to reveal
440 the specific mechanisms involved.
441
442 The presence of oral taxa in the lung microbiota is not unique of disease conditions. It
443 is known that microaspiration, or the aspiration of aerosol droplets originated in the

444 oral cavity, occurs in healthy individuals and can serve as a route for lung colonization
445 of oral commensals[31]. Such an increase of oral bacteria in the lower respiratory tract
446 could be facilitated when critically ill patients –including but not limited to COVID-19–
447 get intubated. Indeed, oral bacteria have been linked to ventilator-associated
448 pneumonia[32,33]. It is yet to be elucidated whether COVID-19 physiopathology favors
449 lung colonization by oral bacteria or if, in contrast, a lung microbiome previously
450 colonized by oral microbes could also contribute to the disease. What is known is that
451 an increase of oral bacteria in the lower respiratory tract can result in an increased
452 inflammatory phenotype, even in healthy subjects[34]
453
454 **Conclusion**
455
456 Overall, this study provides a systematic analysis of potential confounders in COVID-19
457 microbiome studies. We identified that ICU hospitalization and type of oxygen support
458 had large impacts on the upper respiratory tract microbiome and have the potential to
459 confound clinical microbiome studies. Among the different types of oxygen support we
460 reported the largest shifts in microbial community structure between intubated and
461 non-intubated patients. We found that oral microbiota was strongly enriched in the
462 upper and lower respiratory tracts of intubated COVID-19 patients. Further, in the
463 lower respiratory tract, microbes were strongly associated with specific pro-
464 inflammatory immune cells. This information contributes to a collective body of
465 literature on the pathology of COVID-19 and suggests that careful attention be paid to
466 ICU stay and type of oxygen support when evaluating the role of the lung microbiome
467 on COVID-19 disease progression.
468
469 **Methods**
470
471 **Study design and patient cohorts**
472
473 All experimental protocols and data analyses were approved by the Ethics Commission
474 from the UZ Leuven Hospital, under the COntAGIouS observational clinical trial (study
475 number S63381). The study design is compliant with all relevant ethical regulations,

476 including the Declaration of Helsinki and in the GDPR. All participants gave their
477 informed consent to participate in the study.
478
479 A total of 58 patients from the COntAGIouS observational trial were included as our
480 upper respiratory tract cohort. All patients were admitted to the UZ Leuven hospital
481 with a diagnostic of COVID-19. The disease was diagnosed based on a) a positive qRT-
482 PCR test, performed on admission or previously on other hospitals, when patients
483 were transferred from other medical facilities; or b) a chest CT-scan showing alveolar
484 damage and clinical symptoms of the disease. All patients included in the study were
485 admitted to ICU for a variable amount of time. Nasopharyngeal swabs were taken from
486 these patients at different timepoints throughout ICU stay and after ICU discharge,
487 during recovery in ward. A total of 74 swabs were processed for upper respiratory
488 microbiome characterization (Figure 1a).
489
490 To extend our findings from the upper respiratory tract, we also profiled the lower
491 respiratory tract microbiota in a different cohort[22] of 35 patients belonging to the
492 same observational trial and also recruited at UZ Leuven hospital. This cross-sectional
493 cohort is composed by 22 COVID-19 patients and 13 pneumonitis controls with
494 negative qRT-PCR for SARS-CoV-2, with varying disease severity. Previous data from
495 single-cell RNA-sequencing had been collected for this cohort[22]. We reanalyzed this
496 single-cell dataset to profile the lower respiratory tract microbiota in these patients.
497
498
499 **RNA/DNA extraction and sequencing**
500
501 Nucleic acid extraction from the swab samples was performed with AllPrep
502 DNA/RNA/miRNA Universal kit (QIAGEN, catnr. 80224). Briefly, swabs from the
503 potentially infectious samples were inactivated by adding 600µL RLT-plus lysis buffer.
504 To increase bacterial cell lysis efficiency, glass beads and DX reagent (Pathogen Lysis
505 Tubes, QIAGEN, catnr. 19091) were added to the lysis buffer, and samples were
506 disrupted in a FastPrep-24™ instrument with the following program: 1-minute beating
507 at 6.5m/sec, 1-minute incubation at 4°C, 1-minute beating at 6.5m/sec, 1-minute

508 incubation at 4°C. After lysis, the remaining extraction steps followed the
509 recommended protocol from the manufacturer. DNA was eluted in 50µL EB buffer.
510 Amplification of the V4 region of the 16S gene was done with primers 515F and 806R,
511 using single multiplex identifiers and adaptors as previously described[35]. RNA was
512 eluted in 30µL of nuclease-free water and used for SARS-CoV-2 viral load
513 determination in the swabs as well as to measure inflammatory markers and cytokines
514 and to estimate host cell populations via marker gene expression using nCounter. In
515 brief, raw nCounter data were processed using nSolver 4.0 software (Nanostring),
516 sequentially correcting three factors for each individual sample: technical variation
517 between samples (using spiked positive control RNA), background correction (using
518 spiked negative control RNA) and RNA content variation (using 15 housekeeping
519 genes). We have previously validated nCounter digital transcriptomics for
520 simultaneous quantification of host immune and viral transcripts[36], including
521 respiratory viruses in nasopharyngeal aspirates, even with low RNA yield[37–39].
522
523 DNA sequencing was performed on an Illumina MiSeq instrument, generating paired-
524 end reads of 250 base pairs.
525
526 For quality control, reads were demultiplexed with LotuS v1.565[40] and processed
527 following the DADA2 microbiome pipeline using the R packages DADA2[41] and
528 phyloseq[42]. Briefly, reads were filtered and trimmed using the parameters truncQ=11,
529 truncLen=c(130,200), and trimLeft=c(30, 30) and then denoised. After removing
530 chimeras, amplicon sequence variants (ASVs) table was constructed and taxonomy was
531 assigned using the Ribosomal Database Project (RDP) classifier implemented in DADA2
532 (RDP trainset 16/release 11.5). The abundance table was then corrected for copy
533 number, rarefied to even sampling depth, and decontaminated. For decontamination,
534 we used the prevalence-based contaminant identification method in the R package
535 decontam[43].
536
537 **16S statistical analysis**
538

539 All the analyses were performed using R v3.6.0 and the packages vegan[44], phyloseq[42],
540 CoDaSeq[45], DESeq2[46], Biostrings[47], rstatix[48] and DECIPHER[49].
541
542 To analyze the 16S amplicon data, technical replicates were pooled and counts from
543 technical replicates were added. For all the analyses using genus-level agglomerated
544 data, only samples containing more than 10,000 reads assigned at the genus level
545 were used (101 samples in total). Alpha-diversity was analyzed using Shannon's
546 Diversity Index. Comparison of the alpha diversity values across different groups was
547 performed using Wilcoxon signed-rank tests for 2-group comparisons, and Kruskal-
548 Wallis tests for comparisons across multiple groups. In the latter case, pairwise
549 comparisons (when applicable) were performed using Dunn post-hoc tests. To de-
550 confound for the effect of the ICU length, we fitted a quadratic model between the
551 days spent at ICU and the Shannon index using the lm function in R. The residuals of
552 this model were used to test the association with the SARS-CoV-2 viral load.
553
554 Beta diversity analyses were performed using distance-based redundancy analyses
555 (RDA), using Euclidean distances on CLR-transformed data. RDA analyses were
556 performed using the capscale function from vegan. Non-redundant contribution to
557 variation was calculated using the ordiR2step function from vegan, using only the
558 variables that were significant individually in the RDA, and a null model without any
559 explanatory variables. For these analyses, taxa with prevalence lower than 10% were
560 excluded. Metadata variables containing dates, as well as non-informative metadata
561 (containing a single non-NA value or unique for only one patient) were also excluded.
562
563 Differential taxa abundance analyses were performed using DESeq2's likelihood ratio
564 tests and controlling for potential confounders when indicated, including them in the
565 null model. All statistical tests are two-sided, and when multiple tests were applied to
566 the different features (e.g. the differential taxa abundances across two conditions) p-
567 values were corrected for multiple testing using Benjamini-Hochberg's method.
568
569 In order to explore species-level and strain-level diversity, 16S sequences were first
570 clustered into 97% nucleotide diversity operational taxonomic units (OTUs) using the R

571 packages Biostrings and DECIPHER. These OTUs were used to represents the species-
572 level. The number of unique 16S sequences clustered within each OTU were used to
573 represent the number of detectable strains per species. To calculate strain-level
574 diversity per sample, the number of strains of 5 detected OTU species were randomly
575 selected and averaged. This was repeated 1,000x and the average of the all 1,000
576 subsamplings was used as the final strain-level diversity value for each sample, as
577 previously described[50]. To account for uneven sampling assessing diversity differences
578 based on different parameters, we randomly selected and averaged the species- and
579 strain-level diversity of 5 samples per parameter. This was repeated 100x and the
580 subsamplings were used to assess the significant differences between species- and
581 strain-level diversity across the parameters. The average was of all 100 subsamplings
582 was used to as the input for a Pearson's correlation between species- and strain-level
583 diversity.
584
585 **Identification of microbial reads in BAL scRNA-seq data**
586
587 Single-cell data was processed with an in-house pipeline to identify microbial reads.
588 Only read 2, containing the information on the cDNA, was used. Trimmomatic[51] (v0.38)
589 was used to remove trim low quality bases and discard short reads. Additionally,
590 Prinseq++[52] (v1.2) was used to discard reads with low-complexity stretches. Following
591 quality control, reads from human and potential sequencing artifacts (phage phiX174)
592 were mapped with STAR[53] (v2.7.1) and discarded. The remaining reads were mapped
593 against bacterial genomes using a 2-step approach: first, we scanned the reads using
594 mash screen[54] (v2.0) against a custom database of 11685 microbial reference genomes
595 including bacteria, archaea, fungi and viruses. Genomes likely to be present in the
596 analyzed sample (selected using a threshold of at least two shared hashes from mash
597 screen) were selected and reads were pseudoaligned to this subset of genomes using
598 kallisto[55] (v0.44.0). To remove potential artifacts, two additional filters were applied:
599 first, if less than 10 different functions were expressed from a given species, the
600 species was discarded. Second, if one function accounted for more than 95% of the
601 mapped reads of a given species, it was also discarded. These filters were aimed at

602 removing potential artifacts caused by contamination or errors in the genome
603 assemblies used in our database.
604
605 Differences in lower respiratory tract microbial taxa between COVID-19 patients and
606 controls, ICU and ward patients, and invasive and non-invasive ventilation types were
607 calculated using Wilcoxon rank-sum tests on centered-log-ratio (CLR)-transformed
608 data. Prior to CLR data transformation, we filtered the data using the CoDaSeq.filter
609 function, to keep samples with more than 1,000 reads and taxa with a relative
610 abundance above 0.1%. Zeros were imputed using the minimum proportional
611 abundance detected for each taxon. This more lenient approach than the one used for
612 16S data was chosen due to the low number of samples available and the reduced
613 number of bacterial reads identified per sample.
614
615 Bacterial reads were assigned their specific barcodes and UMIs as follows: IDs from the
616 mapped microbial reads were retrieved from the kallisto pseudobam output, and used
617 to retrieve their specific barcodes and UMIs using the raw data files from read 1,
618 assigning each barcode and UMI univocally to a microbial species and function.
619
620 **Direct associations between bacteria and host cells**
621
622 Host single-cell transcriptomics data was obtained from the Seurat[56] object after
623 preprocessing and integrating the samples of the single-cell cohort, as described
624 previously[22]. From the Seurat object, the metadata was extracted, including the
625 information on patient group (COVID-19 or control) and severity of the disease
626 (moderate or critical) as well as cell type and subtype annotation corresponding to
627 each barcode. Enrichment of bacteria detected in patient groups or cell types was
628 calculated using chi-squared tests, with effect sizes determined via the standardized
629 residuals. Significance was assessed via post-hoc tests using the R package
630 chisq.posthoc.test[57].
631
632 For cell types showing an enrichment in associated bacteria, a new Seurat object was
633 created by subsetting the specific cell type. Chi-squared tests were also used to

634 determine enrichment of bacteria-associated cell subtypes. Previous annotations of
635 cell subtypes[22] were used to generate new clusters manually and identify marker
636 genes for these subtypes, using the function findAllMarkers from Seurat. This function
637 was also used to find differentially expressed genes between bacteria-associated and
638 not-bacteria-associated host cells of each subtype. When using this function, reported
639 adjusted p-values are calculated using Bonferroni correction by default.
640
641 **Figure legends**
642
643 **Figure 1. Sample overview and alpha diversity.** a) Longitudinal sampling of patients.
644 Each line represents one patient. Yellow lines span the days spent in ward, while blue
645 lines span the days spent in ICU. Red points mark hospital discharge dates. Crosses
646 indicate the timepoints were swab samples were obtained for microbiome analyses. b)
647 Top 15 most abundant genera in this cohort. Samples with > 10,000 reads assigned to
648 microbial taxa at the genus level were stratified according to the sampling moment:
649 upon admission, throughout the ICU stay or at ICU discharge/during treatment in
650 ward. c) Correlation between the SARS-CoV-2 viral load and Shannon diversity index of
651 all samples. The shaded area surrounding the trend line represents the 95%
652 confidence interval. d) Shannon diversity index of all samples, stratified by the
653 sampling moment: admission, throughout ICU stay or at ICU discharge/during
654 treatment in ward. Boxplots span from the first until the third quartile of the data
655 distribution, and the horizontal line indicates the median value of the data. The
656 whiskers extend from the quartiles until the last data point within 1.5 times the
657 interquartile range, with outliers beyond. Individual data points are also represented.
658
659 **Figure 2. Upper respiratory microbiome covariates in COVID-19.** a) Significant
660 covariates explaining microbiota variation in the upper respiratory tract in this cohort.
661 Individual covariates are listed on the y-axis, their color corresponds to the metadata
662 category they belong to: technical data, disease-related, microbiological tests,
663 comorbidities or host cell populations or gene expression, the latter measured with
664 nCounter (see Methods). Darker colors refer to the individual variance explained by
665 each of these covariates assuming independency, while lighter colors represent the

666 cumulative and non-redundant variance explained by incorporating each variable to a
667 model using a stepwise RDA analysis. The black horizontal line separates those
668 variables that are significant in the non-redundant analysis on top (Patient ID and
669 oxygen support) from the non-significant ones. b) RDA ordination plot showing the
670 first 2 constrained axes. Ordination is constrained by the two significant variables
671 "Patient ID" and "Oxygen support". Samples are depicted as points, whose color
672 indicates the oxygen support type of the patient and whose shape indicates stay at
673 ward or ICU (at the moment of sampling). Axes indicate the variance explained by the
674 first two constrained components of the RDA analysis. c) Species- (left) and strain-level
675 diversity (right) of the samples, stratified by oxygen support type. d) Pearson
676 correlation between average species- and strain- level diversity for each of the oxygen
677 support categories. e) Significant differences among oxygen support types.
678 Differentially abundant taxa between invasive (red) and non-invasive (blue) ventilated
679 sample. Only the top 10 most significant taxa are shown, as determined by their
680 adjusted p-value. Boxplots span from the first until the third quartile of the data
681 distribution, and the horizontal line indicates the median value of the data. The
682 whiskers extend from the quartiles until the last data point within 1.5 times the
683 interquartile range, with outliers beyond. Individual data points are also represented.
684 Gray lines join samples pertaining to the same patient, taken at different time points.
685
686 **Figure 3. Host single cells associated to the lower respiratory tract microbiota.** a)
687 relative proportion of cells from negative and positive COVID-19 patients with (red
688 color) and without (blue) associated bacteria. The p-value of a chi-squared test using
689 the count data is shown on top of the panel. b) Cell types enriched in bacteria-
690 associated cells. Barplots represent the proportion of cell types without ("No") and
691 with ("Yes") bacteria in COVID-19 positive and negative patients. For each patient
692 class, we tested for enrichment of bacteria-associated cells ("Yes") across the different
693 cell types, using the proportions of non-bacteria associated cells ("No") as background.
694 Asterisks mark the cell types with significant enrichment of bacteria. c) Bacterial
695 genera preferentially associating to specific cell types. The heatmaps show the
696 standardized residuals of a chi-squared test including all bacterial genera and the three
697 host cell types enriched in bacteria, for controls (left) and COVID-19 positive patients

698 (right). Taxa with no significant associations with any of the cell types are not shown.
699 Asterisks denote significant positive or negative associations: enrichments are shown
700 in red; depletions are depicted in blue. d) Host cell subtypes associated with bacteria.
701 The heatmap shows the standardized residuals of a chi-squared test including the
702 subtypes of neutrophils, monocytes and monocyte-derived macrophages with
703 associated bacteria, considering cells without bacteria as background. Asterisks denote
704 significant positive or negative associations: enrichments are shown in red; depletions
705 are depicted in blue. e) Marker genes detected for the 5 different subtypes of
706 neutrophils. The heatmap also shows within-group differences between bacteria-
707 associated and bacteria-non-associated cells. f) Myeloid cell functional gene set
708 showing the expression of canonical pro-inflammatory, anti-inflammatory and MHC
709 genes for the two subtypes of myeloid cells significantly associated with bacteria
710 ($CCL2^{hi}$-macrophages and $IL1B^{hi}$-monocytes). The heatmap also shows within group
711 differences between bacteria-associated and bacteria-non-associated cells. Statistically
712 significant differences after multiple testing correction are marked with squares. For
713 b)-d) asterisks denote significance as follows: * = p-value ≤ 0.05; ** = p-value ≤ 0.01;
714 *** = p-value ≤ 0.001; **** = p-value ≤ 0.0001.
715
716
717 **Supplementary Figure Legends**
718
719 **Supplementary Figure 1.** Alpha diversity in the upper respiratory tract. a) Correlation
720 between number of days spent in ICU at the moment of sampling and Shannon
721 diversity index. For samples taken after discharge to ward, the total number of days
722 spent in ICU was used. Spearman correlation and p-value are indicated in the upper
723 right corner of the panel. The shaded area surrounding the trend line represents the
724 95% confidence interval. b) Shannon diversity of the samples, grouped by patient
725 clinical status at the moment of sampling. The p-value of a Kruskal-Wallis test is shown
726 in the upper right corner of the panel. c) Shannon diversity of the samples, grouped by
727 type of oxygen support supplied to the patient at the moment of sampling. The p-value
728 of a Kruskal-Wallis test is shown in the upper right corner of the panel. d) Correlation
729 between the days spent in ICU at the moment of sampling and SARS-CoV-2 viral load

730 of the sample. For samples taken after discharge to ward, the total number of days
731 spent in ICU was used. The shaded area surrounding the trend line represents the 95%
732 confidence interval. e) Correlation between SARS-CoV-2 viral load and Shannon
733 diversity index, after controlling for the time spent in ICU. The residuals of a quadratic
734 fit between the Shannon diversity and the days in ICU were correlated to the SARS-
735 CoV-2 viral loads measured in the samples. Spearman correlation and p-value are
736 indicated in the upper right corner of the panel. The shaded area surrounding the
737 trend line represents the 95% confidence interval. For b) and c), boxplots span from
738 the first until the third quartile of the data distribution, and the horizontal line
739 indicates the median value of the data. The whiskers extend from the quartiles until
740 the last data point within 1.5 times the interquartile range, with outliers beyond.
741 Individual data points are also represented.
742
743 **Supplementary Figure 2.** Differentially abundant taxa between oxygen support types.
744 The 32 taxa whose abundance is significantly different between non-invasive and
745 invasive ventilation are represented. Boxplots span from the first until the third
746 quartile of the data distribution, and the horizontal line indicates the median value of
747 the data. The whiskers extend from the quartiles until the last data point within 1.5
748 times the interquartile range, with outliers beyond. Individual data points are also
749 represented. Gray lines join samples pertaining to the same patient, taken at different
750 time points. An asterisk next to the genera name indicates that the differences
751 between ventilation type are also significant after controlling for patient ID.
752
753 **Supplementary Figure 3.** Absolute microbial read counts in single-cell RNA-seq data
754 from BAL samples. The top 15 species detected in our analyses are depicted. Samples
755 are grouped by disease type (control for non-COVID-19 pneumonia patients, or COVID-
756 19) and hospital stay (ICU or ward).
757
758 **Supplementary Figure 4.** Associations of specific cell types with bacteria, for COVID-19
759 and control samples. The colors represent the strength of the association as the
760 standardized residuals of a Chi-squared test. Red colors indicate a positive association
761 (i.e. enrichment) of bacteria for each cell type. Blue colors indicate a negative

762 association (i.e. depletion) of bacteria for a given cell type. Asterisks denote
763 significance as follows: * = p-value ≤ 0.05; ** = p-value ≤ 0.01; *** = p-value ≤ 0.001;
764 **** = p-value ≤ 0.0001.
765
766 **Author contributions**
767 VLR, ACG, JW, JR designed the study. SJ, TVW, JN, CD, JG, GH, PM collected and
768 processed the BAL samples. PVM and LV collected the clinical data. JW and EW
769 collected the swabs. VLR, ACG and JVW processed the swabs. JVW, MB and SMM
770 determined SARS-CoV-2 viral loads and host gene expression from swabs. DL and JQ
771 generated the single-cell raw data as well as the processed gene-count matrix with
772 annotations of cell types and subtypes. VLR and AGC analyzed the data. VLR, ACG and
773 JR wrote the manuscript. All authors have read and approved the manuscript.
774
775 **Acknowledgments**
776 This study has been supported by funding from the VIB Grand Challenges Program. VLR
777 is supported by an FWO senior postdoctoral fellowship (12V9421N). ACG is supported
778 by an EMBO postdoctoral fellowship (ALTF 349-2019). The Raes lab is supported by KU
779 Leuven, the Rega institute and VIB.
780
781 **Conflict of interest declaration**
782 The authors declare no competing interests.
783
784 **CONTAGIOUS collaborators**
785 Yannick Van Herck, Alexander Wilmer, Michael Casaer, Stephen Rex, Nathalie Lorent,
786 Jona Yserbyt, Dries Testelmans, Karin Thevissen.
787
788 **Data availability**
789 Raw amplicon sequencing data that support the findings of this study have been
790 deposited at the European Genome-phenome Archive (EGA), with accession no. XXX.
791
792 **References**
793
794 1. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients

795 with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**,
796 1054–1062 (2020).
797 2. Grasselli, G. *et al.* Risk Factors Associated with Mortality among Patients with
798 COVID-19 in Intensive Care Units in Lombardy, Italy. *JAMA Intern. Med.* **180**,
799 1345–1355 (2020).
800 3. Mikami, T. *et al.* Risk Factors for Mortality in Patients with COVID-19 in New
801 York City. *J. Gen. Intern. Med.* 1–10 (2020). doi:10.1007/s11606-020-05983-z
802 4. Guo, W. *et al.* Diabetes is a risk factor for the progression and prognosis of
803 <scp>COVID</scp> -19. *Diabetes. Metab. Res. Rev.* **36**, (2020).
804 5. Lighter, J. *et al.* Obesity in Patients Younger Than 60 Years Is a Risk Factor for
805 COVID-19 Hospital Admission. *Clinical infectious diseases : an official publication*
806 *of the Infectious Diseases Society of America* **71**, 896–897 (2020).
807 6. Yu, X. *et al.* SARS-CoV-2 viral load in sputum correlates with risk of COVID-19
808 progression. *Crit. care* **24**, 170 (2020).
809 7. Magleby, R. *et al.* Impact of Severe Acute Respiratory Syndrome Coronavirus 2
810 Viral Load on Risk of Intubation and Mortality Among Hospitalized Patients With
811 Coronavirus Disease 2019. *Clin. Infect. Dis.* (2020). doi:10.1093/cid/ciaa851
812 8. Westblade, L. F. *et al.* SARS-CoV-2 Viral Load Predicts Mortality in Patients with
813 and without Cancer Who Are Hospitalized with COVID-19. *Cancer Cell* (2020).
814 doi:10.1016/j.ccell.2020.09.007
815 9. Coperchini, F., Chiovato, L., Croce, L., Magri, F. & Rotondi, M. The cytokine
816 storm in COVID-19: An overview of the involvement of the
817 chemokine/chemokine-receptor system. *Cytokine and Growth Factor Reviews*
818 **53**, 25–32 (2020).
819 10. Henderson, L. A. *et al.* On the Alert for Cytokine Storm: Immunopathology in
820 COVID-19. *Arthritis and Rheumatology* **72**, 1059–1063 (2020).
821 11. Mehta, P. *et al.* COVID-19: consider cytokine storm syndromes and
822 immunosuppression. *The Lancet* **395**, 1033–1034 (2020).
823 12. Khatiwada, S. & Subedi, A. Lung microbiome and coronavirus disease 2019
824 (COVID-19): Possible link and implications. *Human Microbiome Journal* **17**,
825 100073 (2020).
826 13. Dickson, R. P., Martinez, F. J. & Huffnagle, G. B. The role of the microbiome in
827 exacerbations of chronic lung diseases. *The Lancet* **384**, 691–702 (2014).
828 14. Huffnagle, G. B., Dickson, R. P. & Lukacs, N. W. The respiratory tract microbiome
829 and lung inflammation: A two-way street. *Mucosal Immunology* **10**, 299–306
830 (2017).
831 15. Zhang, H. *et al.* Metatranscriptomic Characterization of COVID-19 Identified A
832 Host Transcriptional Classifier Associated With Immune Signaling. *Clin. Infect.*
833 *Dis.* (2020). doi:10.1093/cid/ciaa663
834 16. Xu, R. *et al.* Temporal dynamics of human respiratory and gut microbiomes
835 during the course of COVID. *medRxiv* 2020.07.21.20158758 (2020).
836 doi:10.1101/2020.07.21.20158758
837 17. Han, Y., Jia, Z., Shi, J., Wang, W. & He, K. The active lung microbiota landscape of
838 COVID-19 patients. *medRxiv* 2020.08.20.20144014 (2020).
839 doi:10.1101/2020.08.20.20144014
840 18. Mostafa, H. H. *et al.* Metagenomic Next-Generation Sequencing of
841 Nasopharyngeal Specimens Collected from Confirmed and Suspect COVID-19

842 Patients Downloaded from. *MBio* **11**, (2020).
843 19. Ho Man, W., de Steenhuijsen Piters, W. A. & Bogaert, D. The microbiota of the
844 respiratory tract: gatekeeper to respiratory health. (2017).
845 doi:10.1038/nrmicro.2017.14
846 20. Whelan, F. J. *et al.* Longitudinal sampling of the lung microbiota in individuals
847 with cystic fibrosis. *PLoS One* **12**, (2017).
848 21. Nolan, T. J. *et al.* Low-pathogenicity Mycoplasma spp. alter human monocyte
849 and macrophage function and are highly prevalent among patients with
850 ventilator-acquired pneumonia. *Thorax* **71**, 594–600 (2016).
851 22. Wauters, E. *et al.* Discriminating Mild from Critical COVID-19 by Innate and
852 Adaptive Immune Single-cell Profiling of Bronchoalveolar Lavages. *Patrick*
853 *Matthys* **9**, 14 (2020).
854 23. Harris, J. R., Balajee, S. A. & Park, B. J. Pneumocystis jirovecii pneumonia:
855 Current knowledge and outstanding public health issues. *Current Fungal*
856 *Infection Reports* **4**, 229–237 (2010).
857 24. Wang, S. *et al.* S100A8/A9 in inflammation. *Frontiers in Immunology* **9**, 1298
858 (2018).
859 25. Coveney, A. P. *et al.* Myeloid-related protein 8 induces self-tolerance and cross-
860 tolerance to bacterial infection via TLR4- and TLR2-mediated signal pathways.
861 *Nat. Publ. Gr.* (2015). doi:10.1038/srep13694
862 26. Holm, J. & Hansen, S. I. Characterization of soluble folate receptors (folate
863 binding proteins) in humans. Biological roles and clinical potentials in infection
864 and malignancy. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1868**,
865 140466 (2020).
866 27. Zakharkina, T. *et al.* The dynamics of the pulmonary microbiome during
867 mechanical ventilation in the intensive care unit and the association with
868 occurrence of pneumonia. doi:10.1136/thoraxjnl-2016-209158
869 28. Schmitt, F. C. F. *et al.* Pulmonary microbiome patterns correlate with the course
870 of disease in patients with sepsis-induced ARDS following major abdominal
871 surgery. (2020). doi:10.1016/j.jhin.2020.04.028
872 29. Shen, Z. *et al.* Genomic Diversity of Severe Acute Respiratory Syndrome-
873 Coronavirus 2 in Patients With Coronavirus Disease 2019. *Clinical infectious*
874 *diseases : an official publication of the Infectious Diseases Society of America* **71**,
875 713–720 (2020).
876 30. Marotz, C. *et al.* Title: Microbial context predicts SARS-CoV-2 prevalence in
877 patients and the hospital built environment. *medRxiv* 2020.11.19.20234229
878 (2020). doi:10.1101/2020.11.19.20234229
879 31. Bassis, C. M. *et al.* Analysis of the upper respiratory tract microbiotas as the
880 source of the lung and gastric microbiotas in healthy individuals. *MBio* **6**, (2015).
881 32. Brennan, M. T. *et al.* The role of oral microbial colonization in ventilator-
882 associated pneumonia. *Oral Surgery, Oral Med. Oral Pathol. Oral Radiol.*
883 *Endodontology* **98**, 665–672 (2004).
884 33. Stonecypher, K. Ventilator-Associated Pneumonia: The Importance of Oral Care
885 in Intubated Adults. *Crit. Care Nurs. Q.* **33**, 339–347 (2010).
886 34. Segal, L. N. *et al.* Enrichment of the lung microbiome with oral taxa is associated
887 with lung inflammation of a Th17 phenotype. *Nat. Microbiol.* **1**, 1–11 (2016).
888 35. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D.

889 Development of a dual-index sequencing strategy and curation pipeline for
890 analyzing amplicon sequence data on the miseq illumina sequencing platform.
891 *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
892 36. Moens, B. *et al.* Simultaneous RNA quantification of human and retroviral
893 genomes reveals intact interferon signaling in HTLV-1-infected CD4+ T cell lines.
894 *Virol. J.* **9**, 171 (2012).
895 37. Fukutani, K. F. *et al.* Pathogen transcriptional profile in nasopharyngeal aspirates
896 of children with acute respiratory tract infection. *J. Clin. Virol.* **69**, 190–196
897 (2015).
898 38. Bouzas, M. L. *et al.* Diagnostic accuracy of digital RNA quantification versus real-
899 time PCR for the detection of respiratory syncytial virus in nasopharyngeal
900 aspirates from children with acute respiratory infection. *J. Clin. Virol.* **106**, 34–40
901 (2018).
902 39. Fukutani, K. F. *et al.* In situ immune signatures and microbial load at the
903 nasopharyngeal interface in children with acute respiratory infection. *Front.*
904 *Microbiol.* **9**, (2018).
905 40. Hildebrand, F., Tadeo, R., Voigt, A. Y., Bork, P. & Raes, J. LotuS: An efficient and
906 user-friendly OTU processing pipeline. *Microbiome* **2**, 30 (2014).
907 41. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina
908 amplicon data. *Nat. Methods* **13**, 581–583 (2016).
909 42. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible
910 Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**,
911 e61217 (2013).
912 43. Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple
913 statistical identification and removal of contaminant sequences in marker-gene
914 and metagenomics data. *Microbiome* **6**, 226 (2018).
915 44. Oksanen, J. *et al.* vegan: Community Ecology Package. (2019).
916 45. Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V. & Egozcue, J. J. It's all relative:
917 analyzing microbiome data as compositions. *Ann. Epidemiol.* **26**, 322–329
918 (2016).
919 46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
920 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
921 47. Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings: Efficient
922 manipulation of biological strings. (2019).
923 48. Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests.
924 (2020).
925 49. Wright, E. S. *Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R*.
926 50. Gregory, A. C. *et al.* Marine DNA Viral Macro- and Microdiversity from Pole to
927 Pole. *Cell* **177**, 1109-1123.e14 (2019).
928 51. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for
929 Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
930 52. Cantu, V. A., Sadural, J. & Edwards, R. PRINSEQ++, a multi-threaded tool for fast
931 1 and efficient quality control and 2 preprocessing of sequencing datasets.
932 (2019). doi:10.7287/peerj.preprints.27553v1
933 53. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–
934 21 (2013).
935 54. Ondov, B. D. *et al.* Mash Screen: high-throughput sequence containment

936 estimation for genome discovery. *Genome Biol.* **20**, 232 (2019).
937 55. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic
938 RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
939 56. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell
940 transcriptomic data across different conditions, technologies, and species. *Nat.*
941 *Biotechnol.* **36**, 411–420 (2018).
942 57. Ebbert, D. chisq.posthoc.test: A Post Hoc Analysis for Pearson's Chi-Squared
943 Test for Count Data. (2019).
944

Figure 1

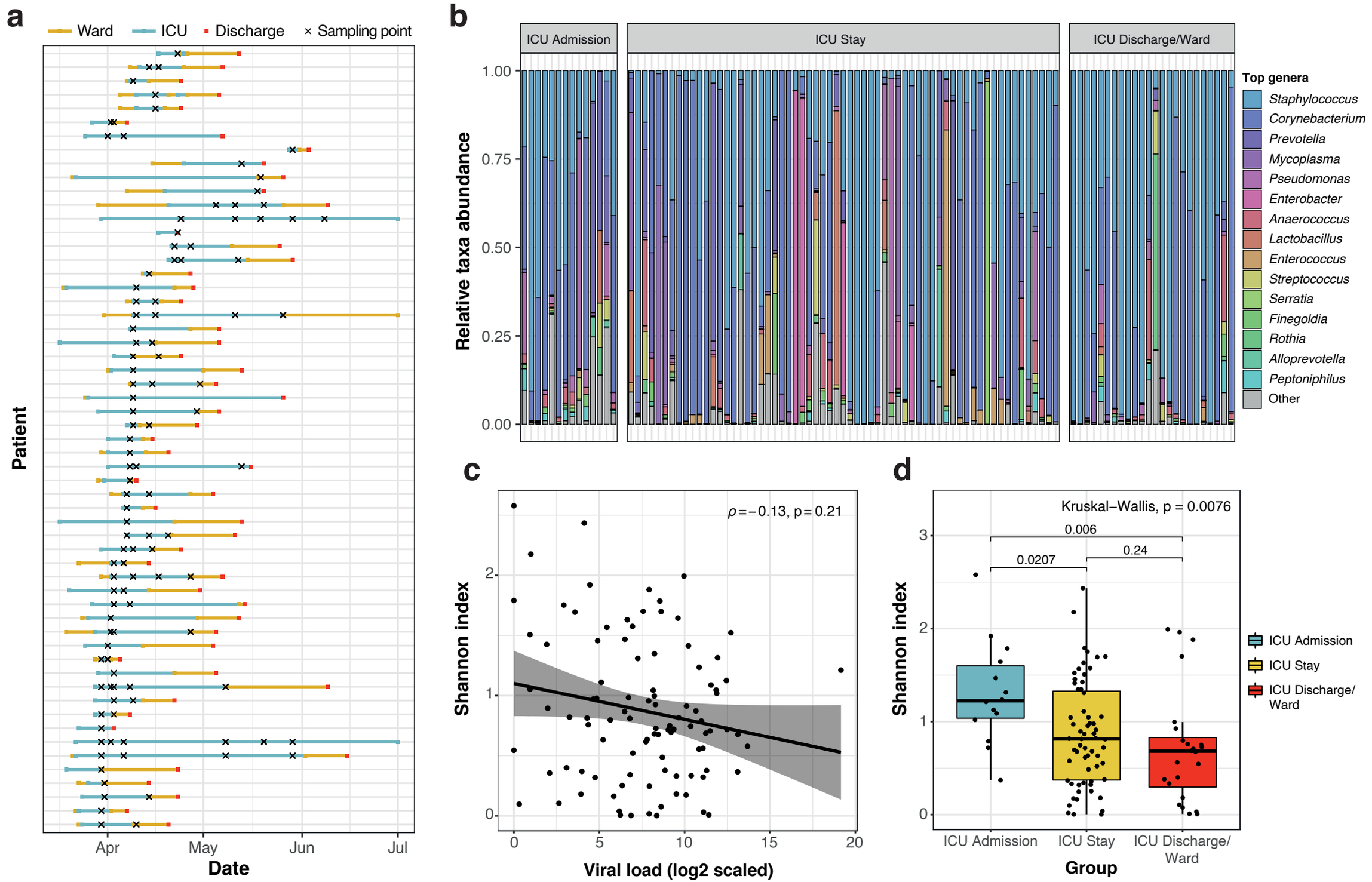

a
Ward
ICU
Discharge
× Sampling point
Patient
Apr
May
Jun
Jul
Date
b
ICU Admission
ICU Stay
ICU Discharge/Ward
Relative taxa abundance
1.00
0.75
0.50
0.25
0.00
Top genera
Staphylococcus
Corynebacterium
Prevotella
Mycoplasma
Pseudomonas
Enterobacter
Anaerococcus
Lactobacillus
Enterococcus
Streptococcus
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

Figure 2

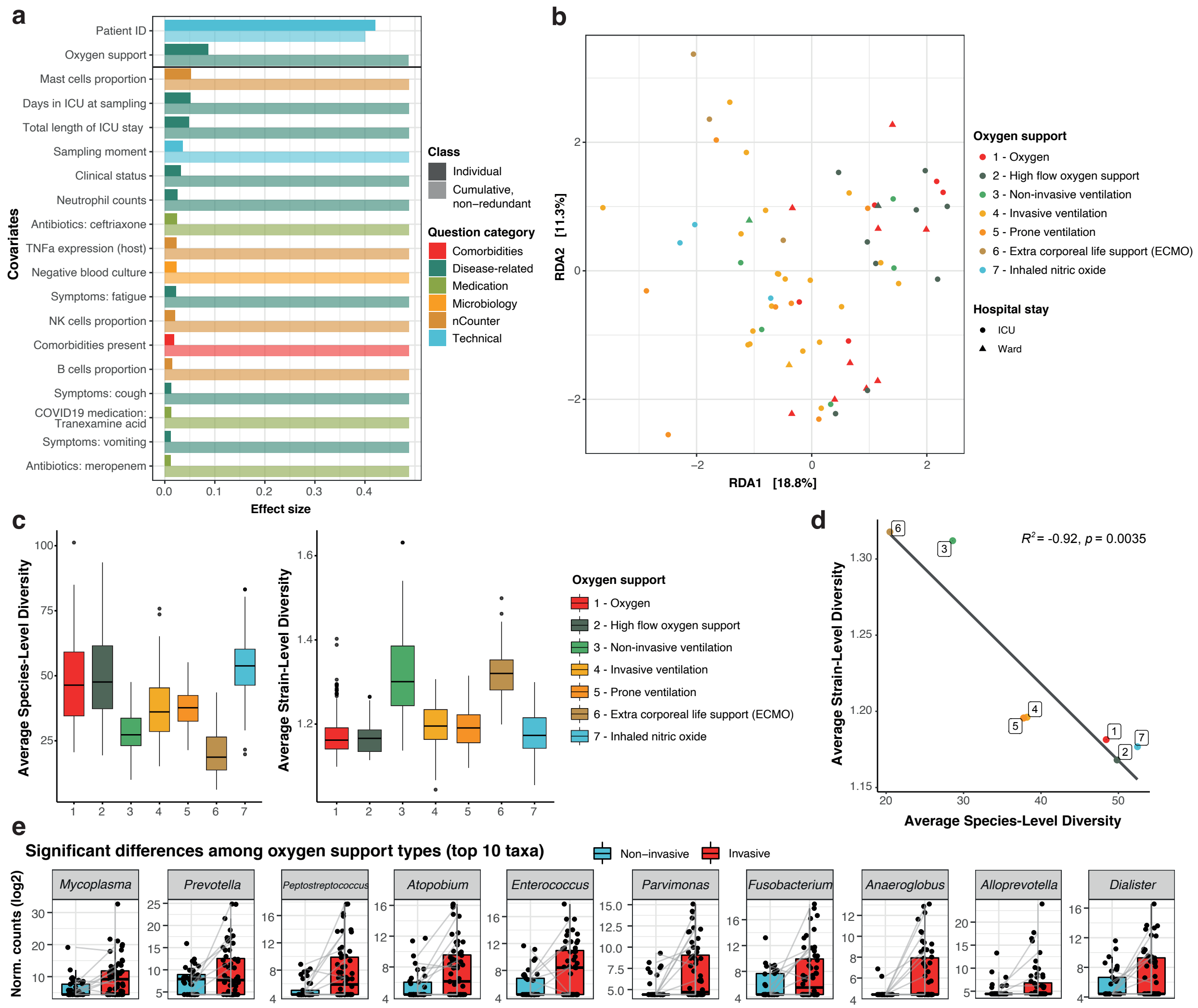
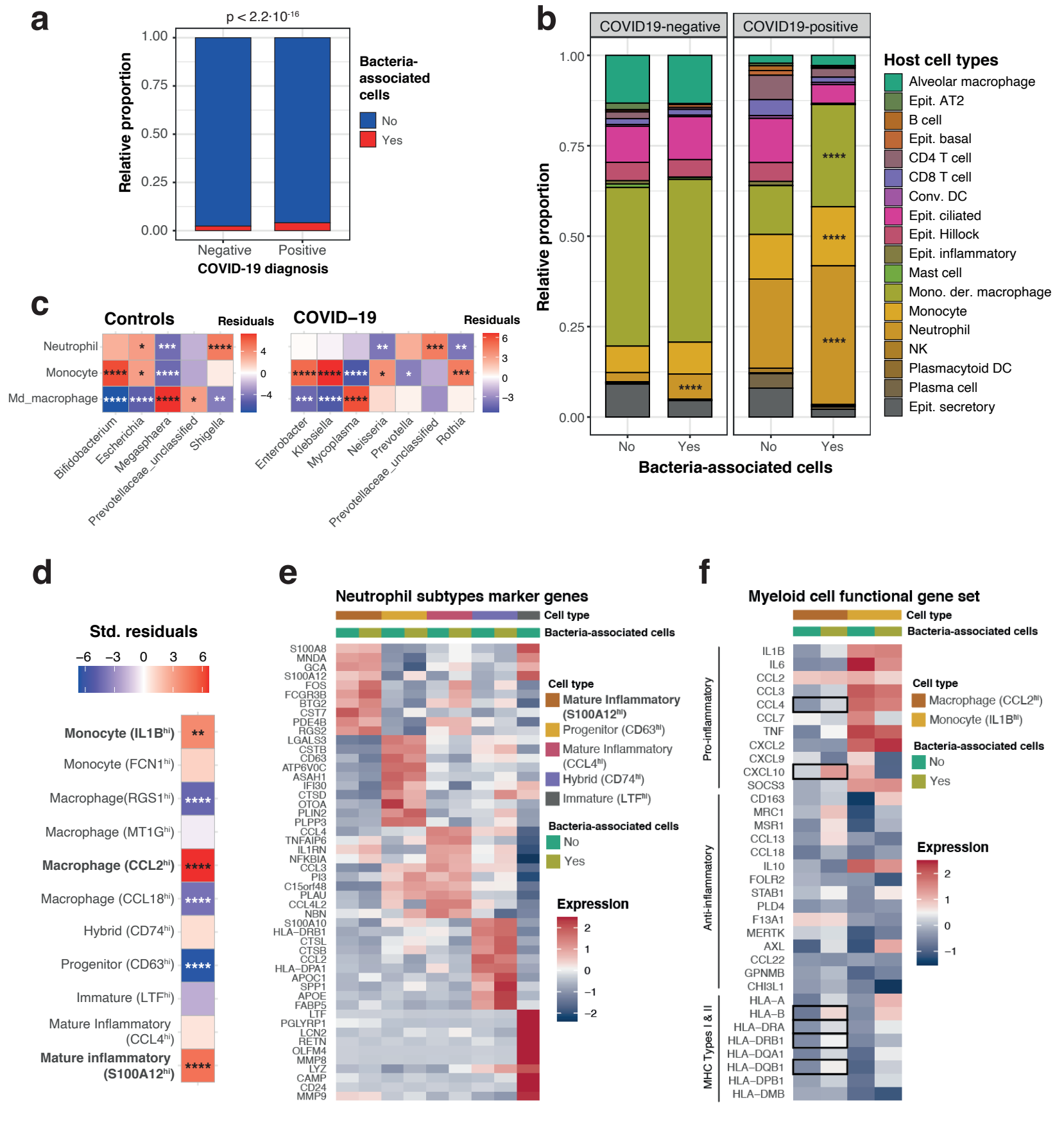

a
Patient ID
Oxygen support
Mast cells proportion
Days in ICU at sampling
Total length of ICU stay
Sampling moment
Clinical status
Neutrophil counts
Antibiotics: ceftriaxone
TNFa expression (host)
Negative blood culture
Symptoms: fatigue
NK cells proportion
Comorbidities present
B cells proportion
Symptoms: cough
COVID19 medication: Tranexamine acid
Symptoms: vomiting
Antibiotics: meropenem
Covariates
0.0
0.1
0.2
0.3
0.4
Effect size
Class
Individual
Cumulative, non–redundant
Question category
Comorbidities
Disease-related
Medication
Microbiology
nCounter
Technical
b
RDA2 [11.3%]
RDA1 [18.8%]
Oxygen support
1 - Oxygen
2 - High flow oxygen support
3 - Non-invasive ventilation
4 - Invasive ventilation
5 - Prone ventilation
6 - Extra corporeal life support (ECMO)
7 - Inhaled nitric oxide
Hospital stay
ICU
Ward
c
Average Species-Level Diversity
Average Strain-Level Diversity
Oxygen support
1 - Oxygen
2 - High flow oxygen support
3 - Non-invasive ventilation
4 - Invasive ventilation
5 - Prone ventilation
6 - Extra corporeal life support (ECMO)
7 - Inhaled nitric oxide
d
R^2 = -0.92, p = 0.0035
Average Strain-Level Diversity
Average Species-Level Diversity
e
Significant differences among oxygen support types (top 10 taxa)
Non–invasive
Invasive
Norm. counts (log2)
Mycoplasma
Prevotella
Peptostreptococcus
Atopobium
Enterococcus
Parvimonas
Fusobacterium
Anaeroglobus
Alloprevotella
Dialister

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

	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