Title: Lung microbiota predict clinical outcomes in critically ill patients

Authors: Robert P. Dickson[1,2,3], Marcus J. Schultz[4,5], Tom van der Poll[6], Laura R. Schouten[4], Nicole R. Falkowski[1], Jenna E. Luth[1], Michael W. Sjoding[1,3,7], Christopher A. Brown[1], Rishi Chanderraj[1,8], Gary B. Huffnagle[1,2], Lieuwe D. J. Bos[4,9]

Affiliations: 1. Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan; 2. Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan; 3. Michigan Center for Integrative Research in Critical Care; Ann Arbor, Michigan; 4. Intensive Care, Amsterdam University Medical Centers, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands. 5. Mahidol University, Bangkok, Thailand: Mahidol-Oxford Tropical Medicine Research Unit (MORU). 6. Centre for experimental and molecular medicine (CEMM), Amsterdam University Medical Center, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands. 7. Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan; 8. Division of Infectious Diseases, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan; 9. Department of Respiratory Medicine, Amsterdam University Medical Centers, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands.

Collaborators: Published on behalf of the BASIC Consortium:

Amsterdam University Medical Centers, location AMC, Amsterdam, The Netherlands: de Beer FM, Bos LD, Claushuis TA, Glas GJ, Horn J, Hoogendijk AJ, van Hooijdonk RT, Huson MA, de Jong MD, Juffermans NP, Lagrand WA, van der Poll T, Scicluna B, Schouten LR, Schultz MJ, van der Sluijs KF, Straat M, van Vught LA, Wieske L, Wiewel MA, Witteveen E.

Corresponding Author Information:

Robert P. Dickson, MD
Pulmonary & Critical Care Medicine
University of Michigan Health System
6220 MSRB III / SPC 5642
1150 W. Medical Center Dr.
Ann Arbor, MI 48109-5642
rodickso@med.umich.edu
(734)277-1487 (mobile)
(734)764-2655 (fax)
(734)936-9368 (lab)
(734)936-5010 (administrator)

Author Contributions: Conception and design: RPD, LDJB. Acquisition of data: MJS, LRS, TVP, LDJB. Processing of specimens and generation of data: NRF, JEL. Analysis and interpretation of data: RPD, MWS, CAB, RC, LDJB. Drafting or revising of manuscript: RPD, MJS, LRS, TVP, NRF, JEL, MWS, CAB, RC, GBH, LDJB. Final approval of manuscript: RPD, MJS, LRS, TVP, NRF, JEL, MWS, CAB, RC, GBH, LDJB.

Funding: Sample and clinical data collection was supported by the Center of Translational Medicine via the MARS consortium. Additional support came from the Academic Medical Center. The authors were supported by the National Institutes for Health (K23HL130641 [RPD], R21AI137669 [RPD], R01HL144599 [RPD]. Additional support was provided by the University of Michigan Center for Integrative Research in Critical Care [RPD].

Running Head: Lung microbiota predict ICU outcomes

Descriptor: 10.9 Pathogen/Host Cell Interactions

Word Count: 3656

At a Glance Commentary

Scientific Knowledge on the Subject

Recent studies have revealed that the lung microbiota of critically ill patients are profoundly altered and are correlated with alveolar and systemic inflammation. Altered lung microbiota may propel and perpetuate alveolar inflammation and injury among critically ill patients. Yet to date, no study has determined whether altered lung microbiota predict disease outcomes in this population.

What This Study Adds to the Field

We here show that among mechanically ventilated critically ill patients, variation in lung microbiota at admission predicts ICU outcomes. Two key features of the lung microbiome, bacterial burden and community composition, predict ventilator-free days. Specifically, increased lung bacterial DNA burden and enrichment of the lung microbiome with gut-associated bacterial taxa (e.g. *Lachnospiraceae* and *Enterobacteriaceae* families) were predictive both of poor ICU outcomes and the clinical diagnosis of ARDS. This correlation between gut-associated bacteria and ARDS validates prior findings, and supports the hypothesis that translocation of gut bacteria to the lungs contributes to the pathogenesis of lung injury. Our results confirm the clinical significance of lung microbiota in critically ill patients. The lung microbiome is an important and underappreciated source of clinical heterogeneity among the critically ill, and represents a novel therapeutic target for the prevention and treatment of lung injury.

Abstract

Rationale: Recent studies have revealed that in critically ill patients, lung microbiota are altered and correlate with alveolar inflammation. The clinical significance of altered lung bacteria in critical illness is unknown.

Objectives: To determine if clinical outcomes of critically ill patients are predicted by features of the lung microbiome at the time of admission.

Methods: We performed a prospective observational cohort study in an intensive care unit (ICU) at a university hospital. Lung microbiota were quantified and characterized using droplet digital PCR and bacterial 16S rRNA gene sequencing. Primary predictors were the bacterial burden, community diversity, and community composition of lung microbiota. The primary outcome was ventilator-free days, determined at 28 days post admission.

Measurements and Main Results: Lungs of 91 critically ill patients were sampled using miniature-bronchoalveolar lavage within 24 hours of ICU admission. Patients with increased bacterial lung bacterial burden had fewer ventilator-free days (HR 0.43, CI 0.21-0.88), which remained significant when controlled for pneumonia and severity of illness. The community composition of lung bacteria predicted ventilator-free days (P=0.003), driven by the presence of gut-associated bacteria (e.g. *Lachnospiraceae* and *Enterobacteriaceae* spp.). Detection of gut-associated bacteria was also associated with the presence of the acute respiratory distress syndrome.

Conclusions: Key features of the lung microbiome (bacterial burden, enrichment with gut-associated bacteria) predict outcomes in critically ill patients. The lung microbiome is an understudied source of clinical variation in critical illness, and represents a novel therapeutic target for the prevention and treatment of acute respiratory failure.

Body

Introduction

In the past decade, advances in culture-independent microbiology have revealed that the lungs, previously considered sterile, harbor complex and dynamic communities of bacteria(1). Lung microbiota are detectable in health(2-4), altered in disease(5, 6), and correlate with variation in airway and alveolar immunity(2, 4, 7). In numerous chronic respiratory diseases, key features of the lung microbiome are predictive of disease outcomes. The burden of lung bacteria (measured by quantification of bacterial DNA) predicts mortality and disease progression in stable patients with idiopathic pulmonary fibrosis(8, 9) and responsiveness to inhaled antibiotics in patients with bronchiectasis(10). The diversity of sputum microbiota predicts mortality in patients with chronic obstructive pulmonary disease(11), and the community composition of respiratory microbiota predicts exacerbations in bronchiectasis(12) and respiratory infections in infants(13).

The lung microbiota of critically ill patients are profoundly altered compared to those of healthy subjects(7, 14-16), and correlate with alveolar and systemic inflammation(7, 15). Specifically, among patients with the acute respiratory distress syndrome (ARDS), the lung microbiome is enriched with gut-associated bacteria(7), and early enrichment of the lung microbiome with gut-associated bacteria (e.g. *Enterobacteriaceae* spp.) is associated with subsequent development of ARDS(15). Altered lung microbiota may propel and perpetuate alveolar inflammation and injury among critically ill patients, but to date no study has determined whether altered lung microbiota predict disease outcomes in this population.

To determine if lung microbiota at ICU admission predict clinical outcomes in critically ill patients, we performed a prospective observational cohort study on critically ill patients receiving mechanical ventilation. The primary outcome was ventilator-free days, adjudicated at

28 days following enrollment. We hypothesized that key features of the lung microbiome (bacterial burden, diversity, and community composition) would predict ICU outcomes, even when controlled for the presence of clinically-appreciated pneumonia.

Methods

Study design

This study was a secondary analysis of specimens collected from patients in the BASIC study (Biomarker Analysis in Septic ICu patients). This study was incorporated in the Molecular Diagnosis and Risk Stratification of Sepsis (MARS) project (NCT01905033) (17-20). The present study was conducted in the ICU of the Academic Medical Center and was approved by the institutional Medical Ethics committee; written informed consent was obtained from the patient representative prior to collection of airway samples via miniature bronchoalveolar lavage (IRB no. NL34294.018.10). Analysis was restricted to baseline specimens and data collected within 24 hours of admission.

Study population

All patients older than 18 years admitted to the ICU with an expected length of stay longer than 24 hours were included in the MARS project. The BASIC study comprised a subset of patients included in the MARS study at the Amsterdam ICU with at least two “systemic inflammatory response syndrome” criteria, who received no antibiotics in the days preceding ICU admission. The current analysis is limited to consecutive patients who were included between September 12, 2011, and November 7, 2013, receiving invasive mechanical ventilation and with informed consent for distal airway sampling. Adjudication of infection was assessed retrospectively using a four-point scale (ascending from none, possible, probable, to definite) using the Centers for Disease Control and Prevention and International Sepsis Forum consensus definitions as previously described(18). ARDS was scored on a daily basis by a team of well-trained clinical

researchers according to the American-European consensus criteria. After the publication of the Berlin definition, all cases were re-evaluated scored according to the new definition, as described previously(21). For the purposes of ARDS vs non-ARDS comparisons, we used adjudication at 24 hours following ICU admission. Severity of illness was quantified using the validated APACHE IV (Acute Physiology and Chronic Health Evaluation IV) (22) and SOFA (Sequential Organ Failure Assessment)(23) models.

Specimen collection and processing

Miniature bronchoalveolar lavage specimens were collected using standard clinical protocol. In short, a 50 cm, 14 Fr tracheal suction catheter was introduced through the orotracheal tube and inserted until significant resistance was encountered. The catheter was then pulled back 1cm and 20 mL 0.9% saline was injected in 10 seconds and immediately aspirated, after which the catheter was removed. Specimens were stored on ice from the time of specimen collection until processing. DNA was extracted, amplified, and sequenced according to previously published protocols(24-26). Sequencing was performed using the Illumina MiSeq platform (San Diego, CA). Bacterial DNA was quantified using a QX200 Droplet Digital PCR System (BioRad, Hercules, CA). Additional details are provided in the online data supplement.

Statistical analysis

As detailed in the online data supplement, we performed microbial ecology analysis using the *vegan* package 2.4-1 and *mvabund* in *R*(27-29) following sequence processing with *mothur*(30, 31). We pre-specified that key features of the microbiome (predictors) would be: 1) bacterial DNA burden (as quantified using droplet digital PCR), 2) bacterial community diversity (as calculated using the Shannon Diversity Index), and 3) community composition. We determined significance in community composition (e.g. mini-BAL specimens vs negative sequencing controls, ARDS vs non-ARDS mini-BAL specimens) using *mvabund* (model-based approach to

analysis of multivariate abundance data). To identify community members driving differences in community composition, we used 1) biplot analysis, 2) rank abundance analysis, and 3) a random forest ensemble learning approach (*randomForest* package in R, version 4.6-14(32)). For random forest, we determined variable importance using 100 forests. The importance parameter was set to TRUE to enable retrieval of importance metrics: mean decrease in accuracy for classification (ARDS vs. non-ARDS) or IncMSE for regression (ventilator-free days). Default settings were utilized for all other parameters. Following model creation, the unscaled feature importance metric was extracted from each forest, assembled into a dataframe, ordered by highest feature importance, and displayed in boxplots of the most important features across the 100 forests. Our primary index of feature importance was Mean Decrease in Accuracy, which ranks predictors by the relative loss in accuracy the occurs when they are removed from the predictive model. We compared means via Student's T test (when normally distributed), the Mann-Whitney U test (when non-Gaussian), and ANOVA with Holm-Sidak's multiple comparisons test as appropriate. Time-to-event analysis was performed using univariate and multivariate Cox proportional hazard models using ventilator-free days (adjudicated 28 days following enrollment) as a primary outcome; multivariate analysis adjusted for age, sex, severity of illness (APACHE IV), diagnosis of ARDS, and the presence of clinically-suspected pneumonia as determined both by the primary clinical service and via post-hoc CDC adjudication criteria. The primary outcome was the proportional hazard ratio for being alive and liberated from mechanical ventilation, as adjudicated 28 days following admission.

Role of the funding source

The funding agencies had no role in the design, conduct, and analysis of the study or in the decision to submit the manuscript for publication.

Results

Study cohort

We obtained mini-BAL specimens from 91 critically ill patients within 24 hours of their ICU admission. The consort diagram is shown in the Appendix Figure E1. Patient demographics and clinical characteristics are reported in **Table 1**.

Table 1. Demographics and Clinical Characteristics of Study Cohort

Characteristic	Study cohort (*n* = 91)
Mean age (SD), *y*	60.7 (15.4)
Male	55 (60)
Admission type	
Medical	67 (74)
Surgical (emergency)	20 (22)
Surgical (elective)	4 (4)
Severity of illness	
Mean SOFA (SD)	7.2 (4.1)
Mean APACHE IV (SD)	82.6 (28.5)
Lung injury	
ARDS at admission	17 (19)
Mean PaO_2:FiO_2 (SD)	262.0 (104.7)
Comorbidities	
Diabetes mellitus	13 (14)
Malignancy	11 (12)
COPD	5 (5)
Immune deficiency	4 (4)
ICU outcomes	
Mean ventilator-free days (SD)	18.5 (10.5)
Mean ICU length of stay, days (SD)	5.6 (4.6)
30-day mortality	27 (30)

SOFA = Sequential Organ Failure Score. APACHE IV = Acute Physiology and Chronic Health Score IV. ARDS = Acute Respiratory Distress Syndrome. COPD = Chronic Obstructive Pulmonary Syndrome. Values are numbers (percentages) unless otherwise indicated.

The distribution of admission diagnoses is reported in **Appendix Table E1**. Bacterial quantification and 16S rRNA gene sequencing was performed on all specimens. Details

regarding adequacy of sequencing and exclusion of specimens are provided in the online supplement.

The microbiota of lung specimens from critically ill patients are distinct from those of background sequencing controls

Low-biomass microbiome studies are vulnerable to contamination due to bacterial DNA present in reagents used in DNA extraction and library preparation(33). Our study used low-volume specimens (mini-BAL) from a low-microbial-biomass body site (the lower respiratory tract), and had their bacterial burden further decreased via a centrifugation step to remove eukaryotic cells(34). Given these concerns, we first asked whether a bacterial signal could be detected in these mini-BAL specimens that was distinct from that of negative controls. We accomplished this by comparing the bacterial DNA in mini-BAL specimens with no-template controls (N = 25), AE buffer specimens (N = 8), sterile water used in DNA extraction (N = 4), extraction control specimens (N = 9), and blank sequencing wells (N = 6).

As shown in **Supplemental Figure 2**, we found clear evidence of distinct bacterial signal in mini-BAL specimens despite their low microbial biomass. Using ultrasensitive quantification of the bacterial 16S gene (via droplet digital PCR), we found significantly more bacterial DNA in mini-BAL specimens than in no-template control specimens ($P < 0.001$, **Supplemental Figure 2A**). The median bacterial DNA burden was 21,246 bacterial gene copies per mL (mean 118,411 copies ± 707,438). We found a wide range of bacterial burden (6,329 - 6,713,947 copies/mL), ranging from comparable to no-template controls to 1000-fold greater than background. Using bacterial community analysis (16S rRNA gene amplicon sequencing), we confirmed that the identity of bacteria detected in mini-BAL specimens was distinct from that of negative control specimens ($P < 0.0001$, *mvabund*). Principal component analysis revealed distinct clustering of mini-BAL specimens apart from negative control specimens

(**Supplemental Figure 2B**), though overlap did occur between some mini-BAL specimens and negative controls. Rank abundance analysis showed clear differences in relative abundance of taxa in negative controls and mini-BAL specimens (**Supplemental Figure 2C**). The dominant taxonomic group in negative controls specimens (OTU008:*Pelomonas*) comprised 25.5% of bacterial sequences in negative controls, but only 2.6% of sequences in mini-BAL specimens.

We thus concluded that while highly variable in their bacterial DNA burden and susceptibility to contamination, mini-BAL specimens contained a distinct bacterial signal from negative control specimens.

Lung microbiota of critically ill patients are altered in patients with ARDS and enriched with gut-associated bacteria (*Enterobacteriaceae* spp.)

We next compared the lung microbiota of critically ill patients with and without ARDS. Prior studies have demonstrated that the lung microbiota of patients with ARDS are altered and enriched with gut-associated bacteria. We compared lung bacterial communities in patients with and without physician-adjudicated ARDS. As shown in **Figure 1**, lung bacterial communities of patients with ARDS differed in the bacterial DNA burden and community composition compared to patients without ARDS.

We first compared ARDS and non-ARDS specimens in their bacterial DNA burden and community diversity. Though highly variable, the bacterial DNA burden in mini-BAL specimens was greater in patients with ARDS than without ARDS (P = 0.014, **Figure 1A**). ARDS specimens did not differ in bacterial community diversity, either measured via the Shannon Diversity Index (P = 0.13) or community richness (P = 0.83) (**Figure 1B**). With both comparisons (bacterial DNA burden and diversity), within-group variation far exceeded across-group differences.

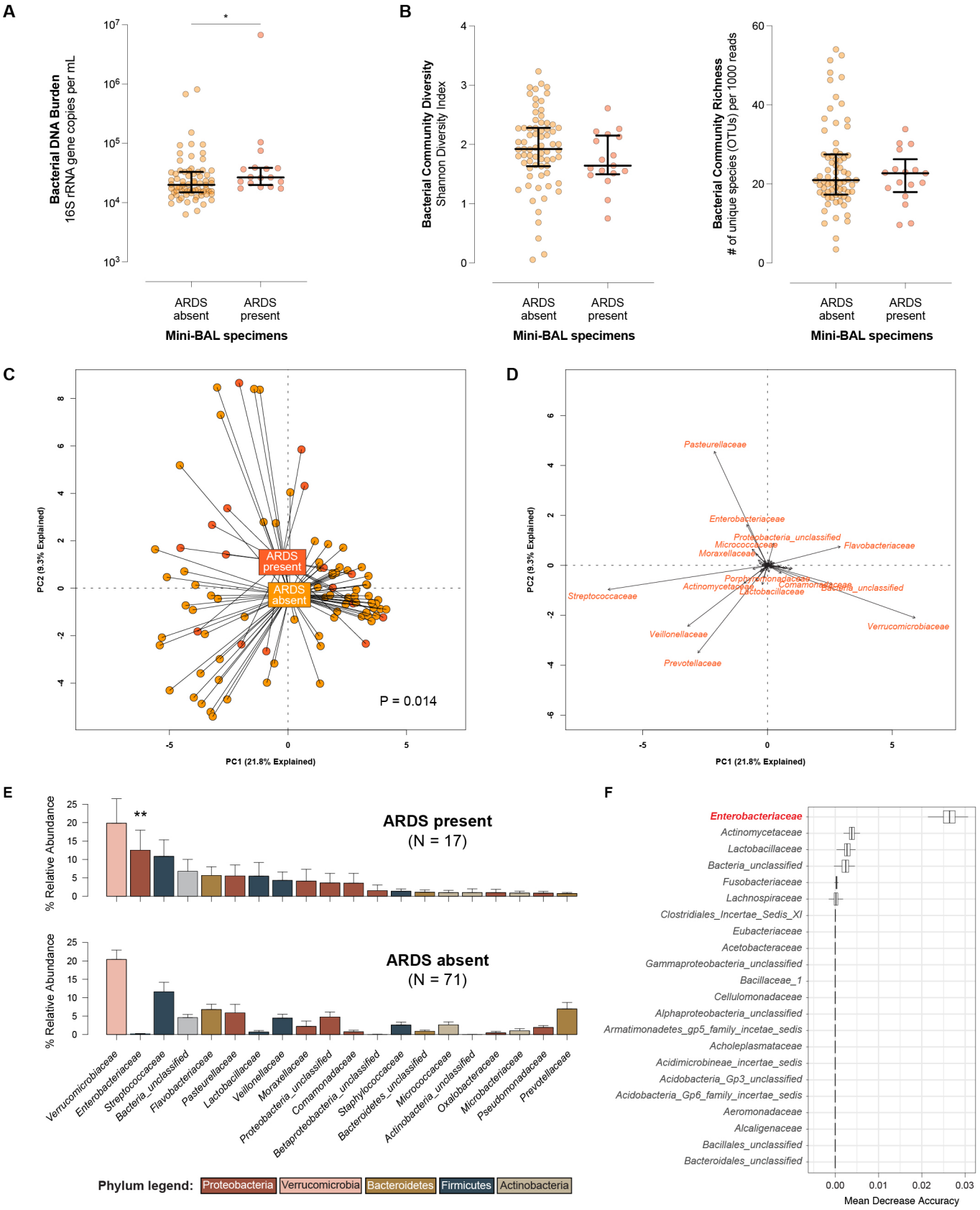


Figure 1. Lung microbiota are altered in patients with the Acute Respiratory Distress Syndrome (ARDS). Compared to patients without ARDS, patients with ARDS had increased lung bacterial burden (A) but no difference in community diversity (B). Principal component analysis of bacterial communities (C) revealed that the community composition of lung bacteria was distinct in specimens from ARDS patients (C), driven by members of the Pasteurellaceae and Enterobacteriaceae families (D). Rank abundance analysis (E) identified taxa enriched in ARDS specimens (e.g. Enterobacteriaceae), and Random Forest analysis (F) confirmed that the Enterobacteriaceae family was the most discriminating taxonomic group between ARDS and non-ARDS specimens. Hypothesis testing performed using Mann-Whitney U test (A), Student's t test (B) and mvabund (C, E). A, B: median and interquartile range.

We next compared the community composition of bacterial communities in ARDS and non-ARDS specimens using complementary approaches. We first visualized communities using principal component analysis (**Figure 1C**). While considerable taxonomic overlap was found across ARDS and non-ARDS specimens, there was a detectable separation of specimens according to ARDS status. This collective difference in community composition was confirmed statistically via *mvabund*, and was robust to taxonomic level of comparison (P = 0.014 at the OTU level of taxonomy, P = 0.013 at the family level, P = 0.003 at the phylum level). We next used biplot analysis to identify specific taxa responsible for this collective difference in community composition (**Figure 1D**). Whereas clusters of non-ARDS specimens were defined by bacterial taxa commonly detected in healthy lungs (*Streptococcaceae* spp., *Veillonellaceae* spp., and *Prevotellaceae* spp.) and taxa detected in negative sequencing control specimens (*Verrucomicrobiaceae* spp., *Flavobacteriaceae* spp.), ARDS specimens were more commonly characterized by *Pasteurellaceae* spp. and *Enterobacteriacaeae* spp..

We then used complementary techniques to identify ARDS-associated bacterial taxa. Using rank abundance visualization (**Figure 1E**), we compared the relative abundance of prominent taxa across ARDS and non-ARDS specimens. While many taxa were common to both group, the *Enterobacteriaceae* family was far more abundant ARDS specimens compared to non-ARDS specimens (12.5% of all bacterial sequences in ARDS specimens compared to 0.18% of all bacterial specimens in non-ARDS specimens). We used unbiased regression-based (*mvabund*) and ensemble-learning (*random forest*) approaches to identify ARDS-enriched taxa. *Mvabund*, which rigorously controls for multiple comparisons, identified the *Enterobacteriaceae* family as enriched in ARDS specimens (P = 0.002). *Random forest* clearly identified the *Enterobacteriaceae* family as the most important taxonomic feature discriminating ARDS from non-ARDS specimens (**Figure 1F**).

We next compared our ARDS-associated *Enterobacteriaceae* taxonomic group with that of an ARDS-associated *Enterobacteriaceae* taxon in a recently published study of mechanically ventilated trauma patients(15). We compared the most prominent *Enterobacteriaceae*-classified OTU in our data set (OTU0005, comprising 61.5% of all *Enterobacteriaceae*-classified sequences) with the ARDS-associated *Enterobacteriaceae* identified by Panzer et al. (OTU2119418). As shown in **Supplemental Figure 3A**, the representative sequence of our study's ARDS-associated *Enterobacteriaceae* OTU was 96% aligned with that of the ARDS-associated *Enterobacteriaceae* OTU identified by Panzer et al., differing in only 3 base pairs. We compared these ARDS-associated OTUs with the taxonomic classifications of closely-aligned sequences from the SILVA ribosomal RNA database. As shown in in **Supplemental Figure 3B**, both OTUs were exclusively identical to *Enterobacteriaceae*-classified taxa, including *Escherichia coli*, *Enterobacter* spp., and *Klebsiella pneumoniae*.

We thus concluded that the lung microbiota of patients with ARDS differ from those of critically ill patients without ARDS, driven by relative enrichment with gut-associated *Enterobacteriaceae* spp..

Lung microbiota are predictive of clinical outcomes in critically ill patients

We next asked if key features of the lung microbiome (*bacterial burden*, *diversity*, and *community composition*) predict clinical outcomes in critically ill patients. Our primary outcome was ventilator-free days measured at 28 days following admission.

We first asked if *bacterial burden* of mini-BAL specimens (quantified using ddPCR of the 16S rRNA gene) predicted ICU outcomes (**Table 2**). Using univariate analysis, we found that increased baseline lung bacterial DNA burden predicted fewer ventilator-free days, either when

analyzed continuously (hazard ratio 0.43, confidence interval 0.21 - 0.88, P = 0.022) or when comparing tertiles defined by total lung bacterial DNA burden. In other words, for each additional 10-fold increase in lung bacterial DNA, the Hazard Ratio for favorable outcome (liberation from mechanical ventilation) was 0.43. As shown in **Figure 2**, the tertile of patients with the highest baseline lung bacterial DNA burden were less likely to be extubated and alive at 7, 14, 21, and 28 days compared to patients with low bacterial DNA burden (HR 0.45, confidence interval 0.25 - 0.81, P = 0.008).

Table 2. Predictors of Ventilator-Free Days in Mechanically Ventilated Critically Ill Patients

Predictor	*Univariate*		*Multivariate*	
	Hazard ratio (CI)	***P* value**	**Hazard ratio (CI)**	***P* value**
Lung bacterial DNA burden (continuous)	0.43 (0.21 - 0.88)	0.022	0.40 (0.18 - 0.86)	0.019
Lung bacterial DNA burden: middle tertile*	0.87 (0.50 - 1.51)	0.62		
Lung bacterial DNA burden: highest tertile*	0.45 (0.25 - 0.81)	0.008		
Shannon diversity index†	1.27 (0.87 - 1.86)	0.21		
Age (years)	0.99 (0.98 - 1.01)	0.35	1.01 (0.99 - 1.03)	0.32
Gender (male)	1.26 (0.78 - 2.03)	0.35	0.90 (0.54 - 1.49)	0.68
SOFA	0.95 (0.90 - 1.01)	0.10		
APACHE IV	0.98 (0.98 - 0.99)	<0.001	0.98 (0.97 - 0.99)	<0.001
Suspected pneumonia	1.01 (0.60 - 1.70)	0.96	0.90 (0.53 - 1.55)	0.71
Pneumonia (post-hoc, CDC criteria)	0.48 (0.18 - 1.33)	0.16		
ARDS	0.51 (0.27 - 0.98)	0.044	0.61 (0.31 - 1.21)	0.16

CI = Confidence interval (95%). SOFA = Sequential Organ Failure Score. APACHE IV = Acute Physiology and Chronic Health Score IV. ARDS = Acute Respiratory Distress Syndrome.
*Versus lowest tertile.
†Shannon diversity index, per 1-unit increase

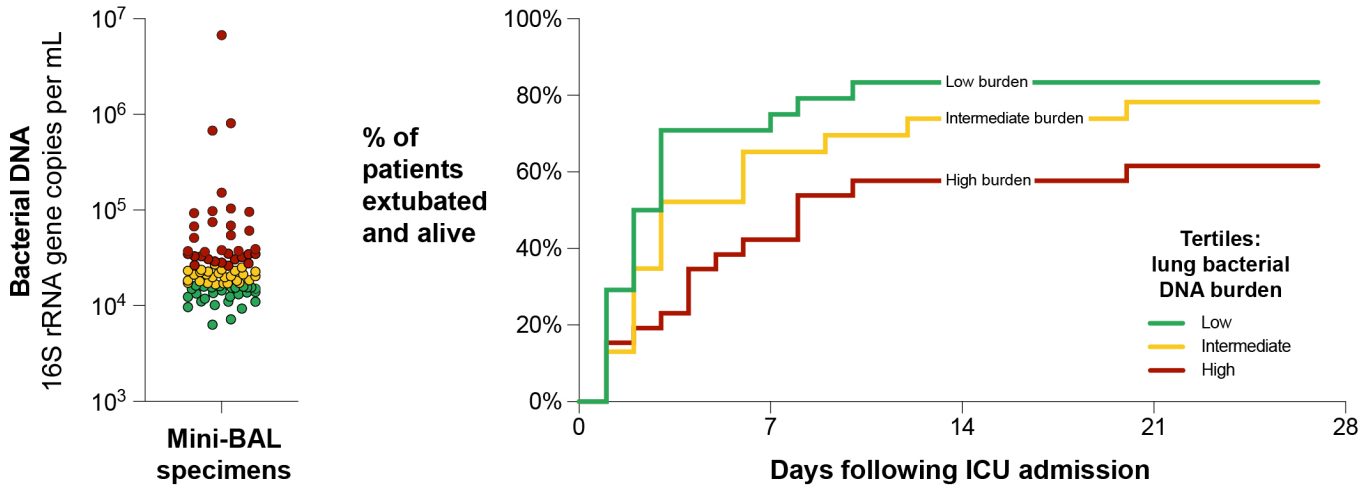


Figure 2. Lung microbiota predict 28-day outcomes in mechanically-ventilated critically ill patients. In critically ill patients receiving mechanical ventilation, the burden of bacterial DNA detected in mini-BAL specimens was predictive of total ventilator-free days. Patients with high lung burdens of bacterial DNA were less likely to be extubated and alive than patients with low bacterial DNA burden (P = 0.008). Hypothesis testing performed using univariate Cox proportional hazard modeling.

Pneumonia is common among mechanically ventilated patients, and a potential source of confounding in lung microbiome studies. In our study cohort, 27 patients (30%) had suspected pneumonia either at admission or within 48 hours; of these, 7 patients met CDC criteria for probable or confirmed pneumonia via post-hoc chart review. When we controlled our outcome analysis of lung bacterial DNA burden for the presence of clinically suspected pneumonia, it did not meaningfully change either the hazard ratio or significance of the model (HR = 0.43, P = 0.021). Similarly, controlling for post-hoc adjudicated pneumonia (using CDC criteria) also did not influence the predictive power of lung bacterial DNA burden (HR = 0.43, P = 0.019). We thus concluded that lung bacterial burden predicts poor outcomes in mechanically ventilated critically ill patients, even when controlled for the presence of suspected or confirmed pneumonia.

We then performed multivariate analysis to determine whether lung bacterial DNA burden is independently predictive of poor outcomes. The relationship between increased lung bacterial DNA burden and fewer ventilator-free days remained significant when controlled for age,

gender, severity of illness (APACHE IV score), the presence of suspected pneumonia, and the presence of ARDS (HR 0.40, confidence interval 0.18 - 0.86, P = 0.019). We thus concluded that lung bacterial DNA burden is an independent predictor of poor outcomes in critically ill patients.

We next asked if *bacterial diversity* of lung bacteria predicts ICU outcomes (**Figure 3A**). Bacterial diversity of lung bacteria (as measured by the Shannon Diversity Index) did not significantly predict ICU outcomes (P = 0.22). While the most favorable ICU outcomes were observed among patients with high baseline lung bacterial diversity, followed in stepwise manner by patients with intermediate and low diversity, this difference in tertiles was not statistically significant. Other indices of lung bacterial diversity (community richness, community dominance) were also not significantly predictive of ICU outcomes (ventilator-free days, P > 0.05 for all comparisons, **Appendix Table E2**).

Finally, we asked if the *community composition* of lung bacteria is predictive of ICU outcomes. We compared patient ventilator-free days with lung bacterial community structure using *mvabund* (model-based approach to analysis of multivariate abundance data). The overall community composition of baseline lung microbiota was significantly predictive of patient ventilator-free days (P = 0.003 at the OTU level of taxonomy, P = 0.004 at the family level). Using *random forest* to identify taxa associated with poor outcomes, we identified the gut-associated *Lachnospiraceae* and *Enterobacteriaceae* families as the taxa most strongly predictive of fewer ventilator-free days (**Figure 3B**). We then tested the hypotheses that detection of these taxa predicts poor outcomes. As shown in **Figure 3B**, detection of the *Lachnospiracaeae* family was significantly predictive of worse ICU outcomes (P = 0.020). The relationship between *Enterobacteriaceae* detection and ventilator-free days was not significant (P = 0.12).

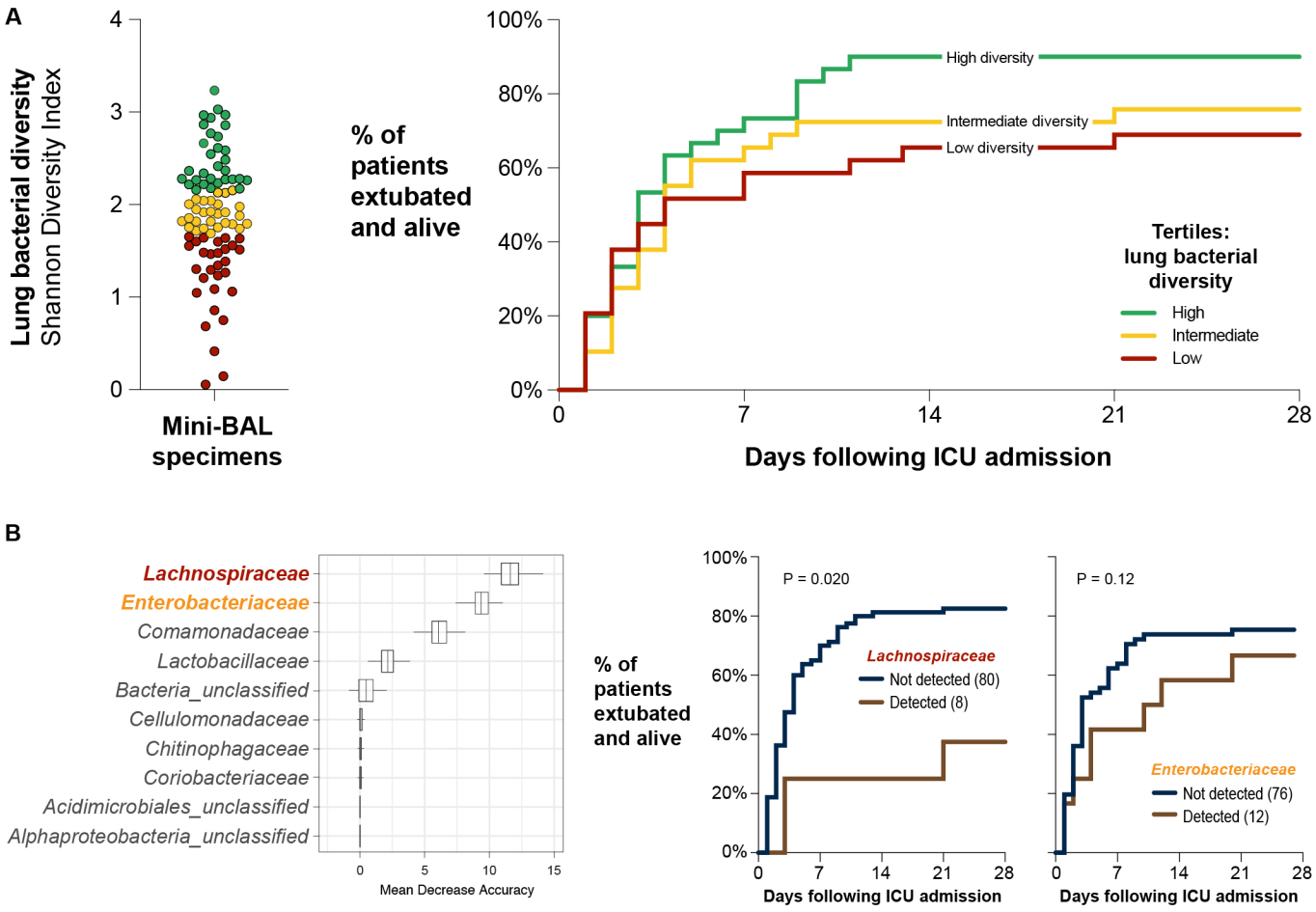


Figure 3. Lung microbiota and 28-day outcomes in mechanically-ventilated critically ill patients. Community diversity of lung bacteria was highly variable among patients, and did not significantly predict ventilator-free days (A). Community composition of lung bacteria was significantly predictive of ventilator-free days (P = 0.003, mvabund). Random forest identified the gut-associated *Lachnospiracea* and *Enterobacteriaceae* families as the strongest predictors of ventilator-free days (B). Hypothesis testing performed using Cox proportional hazard model.

We thus concluded that in the lung microbiota of critically ill patients, poor ICU outcomes are predicted both by increased bacterial burden and by community composition (specifically, enrichment with gut-associated taxa).

Discussion

The core finding of this study is that among mechanically ventilated critically ill patients, variation in lung microbiota at admission predicts ICU outcomes. Two key features of the lung microbiome, bacterial burden and community composition, predicted ventilator-free days. Specifically, increased lung bacterial DNA burden and enrichment of the lung microbiome with

316 gut-associated bacterial taxa (e.g. *Lachnospiraceae* and *Enterobacteriaceae* families) were
317 predictive of poor ICU outcomes and the clinical diagnosis of ARDS.
318
319 Our study is the first to demonstrate that variation in lung microbiota is predictive of clinical
320 outcomes in critically ill patients. This key finding is compatible with broader (non-ICU) lung
321 microbiome studies, which have found that lung microbiota are predictive of disease outcomes
322 in idiopathic pulmonary fibrosis(8, 9), chronic obstructive pulmonary disease(11),
323 bronchiectasis(10, 12), and infants susceptible to respiratory infections(13). While these
324 findings, robust across disease states, confirm that the lung microbiome is a risk factor for
325 disease progression, a crucial and unanswered question is whether lung microbiota are a
326 *modifiable* risk factor. Animal studies using germ-free and antibiotic-exposed mice suggest that
327 manipulation of the microbiome does influence host susceptibility to lung inflammation, injury,
328 and mortality(9, 35). Yet the precise role of gut and lung microbiota in the pathogenesis of acute
329 and chronic lung injury has yet to be elucidated. Future studies should interrogate whether the
330 microbiome's role in lung disease is more attributable to remote (gut-lung) or local (lung-lung)
331 host-microbiome interactions(2).
332
333 Our findings both validate several recent studies and provide new insight into the importance of
334 the lung microbiome in critical illness. We have previously reported that the lung microbiome is
335 enriched with gut-associated bacteria in ARDS and correlated with the severity of systemic and
336 alveolar inflammation(7). In a subsequent study of mechanically ventilated trauma patients, the
337 presence of gut-associated bacteria in endotracheal aspirates (*Enterobacteriaceae* spp.) was
338 associated with ARDS onset(15). In our current study, we found that the lung microbiota of
339 patients with ARDS was distinct from patients without ARDS, again driven by the presence of
340 gut-associated bacteria (*Enterobacteriaceae* spp.). Indeed, the bacterial taxon most strongly
341 correlated with ARDS status in our study (OTU0005:*Enterobacteriaceae*) was nearly identical to

that of the ARDS-associated bacterial taxon found by Panzer and colleagues(15). These multiple findings, now robust across cohorts, sequencing platforms, laboratories, and continents, all provide indirect support for the hypothesis that gut-lung translocation of bacteria contributes to the pathogenesis of lung injury in critically ill patients.

Importantly, our core findings remained significant when controlled for the clinical suspicion or post-hoc adjudication of pneumonia. While this may seem paradoxical (increased lung bacterial burden predicts poor outcomes, and pneumonia is a condition of high lung bacterial burden), recent culture-independent studies have revealed both the complexity of lung bacterial communities in mechanically ventilated patients(14, 16, 36) and the inadequacy of our conventional understanding of pneumonia(37). The lack of concordance between our molecular characterization of lung bacteria and clinical assessment of pneumonia likely reflects several key issues in the microbiology of injured lungs: 1) clinical adjudication of pneumonia, especially in mechanically ventilated patients, is imprecise and unreliable(38), 2) a dichotomous adjudication of pneumonia is too simplistic and reductionistic to meaningfully describe the complex ecologic spectrum of respiratory microbiota, and 3) the lung microbiome may play a role in the pathogenesis of disease processes not classically considered infectious (e.g. perpetuating inflammation and injury in ARDS). Our results highlight the need for improved molecular diagnostics to provide clinicians with a more accurate and comprehensive assessment of lung microbiota, as well as a more refined ecologic understanding of respiratory infections in critically ill patients.

Our study has several limitations that should prompt further validation and study. While we detected a distinct bacterial signal in our specimens, the bacterial biomass in these cell-free mini-BAL specimens was low, and in many specimens overlapped with background “sequencing noise.” Future studies, using larger volumes of whole BAL, may find stronger bacterial signal.

Our mini-BAL sampling approach was non-directional; thus the anatomic site of sampling was not standardized across patients. While our findings remained significant when controlled for important clinical confounders, we could not control for all potential exposures (e.g. ICU antibiotic exposure or pre-ICU medications), and residual confounding is likely. Finally, while our findings provide indirect support for the hypothesis of gut-lung translocation contributing to lung injury in critically ill patients, our lack of paired gut specimens precludes our determining whether gut-associated taxa (e.g. *Enterobacteriaceae* and *Lachnospiracaeae* spp.) were derived from the lower gastrointestinal tract or via another route (e.g. aspiration of altered pharyngeal microbiota). Future prospective studies of critically ill patients, in addition to sampling the lower respiratory tract, should collect time-matched specimens from the lower and upper gastrointestinal tract.

In conclusion, in this prospective observational cohort study of mechanically ventilated critically ill patients, variation in baseline lung microbiota predicted ICU outcomes. Increased lung bacterial burden and lung enrichment with gut-associated bacteria were predictive of worse outcomes. The lung microbiome is an important and understudied source of variation among critically ill patients, and may represent a novel therapeutic target for the prevention and treatment of lung injury.

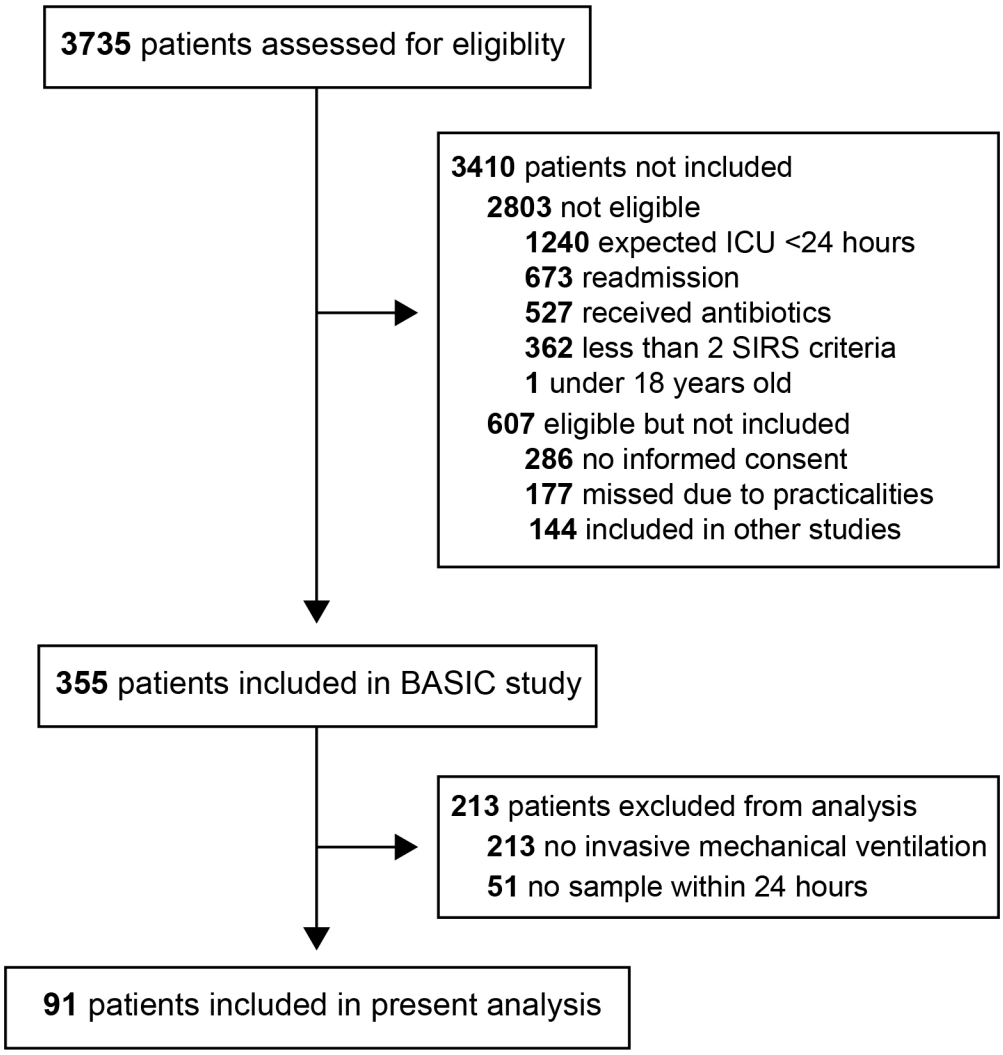
Data availability

Sequences are available via the NCBI Sequence Read Archive (accession number PRJNA553560). OTU tables, taxonomy classification tables, and metadata tables are available at https://github.com/dicksonlunglab/MARS_lung_microbiome.

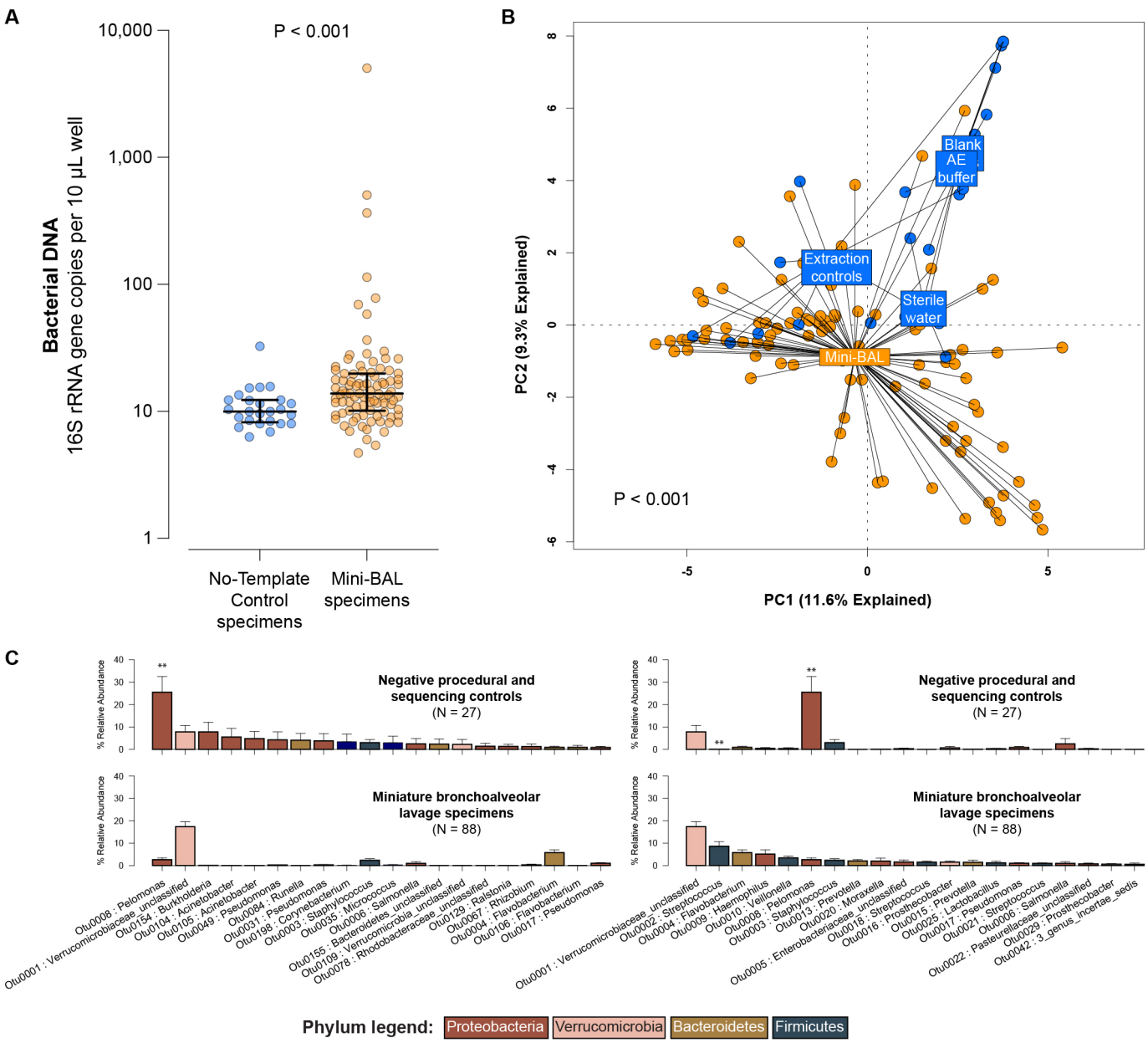
Acknowledgements

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health**.** The authors thank Carolyn Calfee, Ariane Panzer, and Susan Lynch for sharing the representative sequence of OTU2119418:*Enterobacteriaceae*.

396



Supplemental Figure 1. CONSORT flow diagram of patient screening, inclusion, and analysis. The BASIC study (Biomarker Analysis in Septic Intensive Care patients) was a prospective, observational cohort study, of which the current study is a sub-analysis.



Supplemental Figure 2. Evidence of distinct bacterial signal in mini-bronchoalveolar lavage specimens compared to negative control specimens. Bacterial DNA in miniature-bronchoalveolar lavage (mini-BAL) specimens was quantified and identified using droplet digital PCR (A) and 16S rRNA gene amplicon sequencing (B, C), respectively. (A) The bacterial burden of mini-BAL specimens, while highly variable, was significantly greater than that of negative control specimens. (B) The community composition of bacteria identified in mini-BAL specimens was significantly distinct from that of bacteria detected in negative control specimens. (C) Rank abundance comparison of prominent taxa detected in negative controls (left) and mini-BAL (right) specimens. For each comparison, the twenty most abundant taxa in each group are displayed in decreasing order of relative abundance (mean ± SD). Asterisks indicated taxa idendified as significantly distinct across groups. Hypothesis testing performed using Mann-Whitney U test (A) and *mvabund* (B,C). A: median and interquartile range. B: mean and standard deviation.

Supplemental Figure 3. (A) In our study, the presence of ARDS was associated with enrichment of an operational taxonomic unit (OTU) classified as *Enterobacteriaceae* (Otu0005, orange). This sequence of bacterial DNA was 96% aligned with an ARDS-associated *Enterobacteriaceae* OTU from a prior study of the lung microbiome in critically ill patients (Panzer *AJRCCM* 2018) (Otu2119418, light blue). (B) For each *Enterobacteriaceae* OTU, we identified the 20 most closely-aligned sequences identified via the SILVA database project and compared these OTUs with their neighbors, each other, and other prominent taxa from our data set (classified as *Streptococcus*, *Staphylococcus*, and *Haemophilus*).

Table E1. Documented Admission Diagnoses of Study Cohort

Diagnosis	Study cohort (*n* = 91)
Cardiac arrest	19 (21)
Cerebral vascular accident (ischemic or hemorrhagic)	9 (10)
Pneumonia	7 (8)
Congestive heart failure/cardiomyopathy	6 (7)
Aortic aneurysm (rupture or dissection)	5 (5)
Sepsis	5 (5)
Meningitis	4 (4)
Pulmonary embolism	3 (3)
Bowel perforation	3 (3)
Gastrointestinal hemorrhage	2 (2)
Other	28 (31)

"Other" includes: arrhythmia, cholangitis, elective surgery, hemoptysis, hypoglycemia, obtundation, pyelonephritis, rhabdomyolysis, seizures, soft tissue infection, trauma.

Table E2. Diversity Indices as Predictors of Ventilator-Free Days in Mechanically Ventilated Critically Ill Patients

Predictor	*Univariate*	
	Hazard ratio (CI)	*P* value
Shannon diversity index*	1.27 (0.87 - 1.86)	0.21
Community richness†	1.01 (0.98 - 1.00)	0.09
Community dominance‡	1.02 (1.00 - 1.04)	0.10

CI = Confidence interval (95%).
* per 1-unit increase.
† Unique OTUs per 1000 sequences, per OTU
‡ Relative abundance of most dominant OTU, per % relative abundance

References

1. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The microbiome and the respiratory tract. *Annu Rev Physiol* 2016; 78: 481-504.

2. Dickson RP, Erb-Downward JR, Falkowski NR, Hunter EM, Ashley SL, Huffnagle GB. The lung microbiota of healthy mice are highly variable, cluster by environment, and reflect variation in baseline lung innate immunity. *Am J Respir Crit Care Med* 2018; 198: 497-508.

3. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, Curtis JL. Bacterial topography of the healthy human lower respiratory tract. *MBio* 2017; 8.

4. Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG, Li Y, Shen N, Ghedin E, Morris A, Diaz P, Huang L, Wikoff WR, Ubeda C, Artacho A, Rom WN, Sterman DH, Collman RG, Blaser MJ, Weiden MD. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat Microbiol* 2016; 1: 16031.

5. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet* 2014; 384: 691-702.

6. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt MF, Cookson WO. Disordered microbial communities in asthmatic airways. *PLoS One* 2010; 5: e8578.

7. Dickson RP, Singer BH, Newstead MW, Falkowski NR, Erb-Downward JR, Standiford TJ, Huffnagle GB. Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nat Microbiol* 2016; 1: 16113.

8. Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, Murphy E, Johnston SL, Schwartz DA, Wells AU, Cookson WO, Maher TM, Moffatt MF. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 190: 906-913.

9. O'Dwyer DN, Ashley SL, Gurczynski SJ, Xia M, Wilke C, Falkowski NR, Norman KC, Arnold KB, Huffnagle GB, Salisbury ML, Han MK, Flaherty KR, White ES, Martinez FJ, Erb-Downward JR, Murray S, Moore BB, Dickson RP. Lung microbiota contribute to pulmonary inflammation and disease progression in pulmonary fibrosis. *Am J Respir Crit Care Med* 2019.

10. Sibila O, Laserna E, Shoemark A, Keir HR, Finch S, Rodrigo-Troyano A, Perea L, Lonergan M, Goeminne PC, Chalmers JD. Airway bacterial load and inhaled antibiotic response in bronchiectasis. *Am J Respir Crit Care Med* 2019; 200: 33-41.

11. Leitao Filho FS, Alotaibi NM, Ngan D, Tam S, Yang J, Hollander Z, Chen V, FitzGerald JM, Nislow C, Leung JM, Man SFP, Sin DD. Sputum Microbiome Is Associated with 1-Year Mortality after Chronic Obstructive Pulmonary Disease Hospitalizations. *Am J Respir Crit Care Med* 2019; 199: 1205-1213.

12. Rogers GB, Zain NM, Bruce KD, Burr LD, Chen AC, Rivett DW, McGuckin MA, Serisier DJ. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Ann Am Thorac Soc* 2014; 11: 496-503.

13. Bosch A, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Biesbroek G, Kool J, Pernet P, de Groot PCM, Eijkemans MJC, Keijser BJF, Sanders EAM, Bogaert D. Maturation of the infant respiratory microbiota, environmental drivers, and health consequences. A prospective cohort study. *Am J Respir Crit Care Med* 2017; 196: 1582-1590.

14. Kelly BJ, Imai I, Bittinger K, Laughlin A, Fuchs BD, Bushman FD, Collman RG. Composition and dynamics of the respiratory tract microbiome in intubated patients. *Microbiome* 2016; 4: 7.

15. Panzer AR, Lynch SV, Langelier C, Christie JD, McCauley K, Nelson M, Cheung CK, Benowitz NL, Cohen MJ, Calfee CS. Lung microbiota is related to smoking status and to development of ARDS in critically ill trauma patients. *Am J Respir Crit Care Med* 2017; 197: 621-631.

16. Zakharkina T, Martin-Loeches I, Matamoros S, Povoa P, Torres A, Kastelijn JB, Hofstra JJ, de Wever B, de Jong M, Schultz MJ, Sterk PJ, Artigas A, Bos LDJ. The dynamics of the pulmonary microbiome during mechanical ventilation in the intensive care unit and the association with occurrence of pneumonia. *Thorax* 2017; 72: 803-810.

17. Scicluna BP, Klein Klouwenberg PM, van Vught LA, Wiewel MA, Ong DS, Zwinderman AH, Franitza M, Toliat MR, Nurnberg P, Hoogendijk AJ, Horn J, Cremer OL, Schultz MJ, Bonten MJ, van der Poll T. A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. *Am J Respir Crit Care Med* 2015; 192: 826-835.

18. Klein Klouwenberg PM, Ong DS, Bos LD, de Beer FM, van Hooijdonk RT, Huson MA, Straat M, van Vught LA, Wieske L, Horn J, Schultz MJ, van der Poll T, Bonten MJ, Cremer OL. Interobserver agreement of Centers for Disease Control and Prevention criteria for classifying infections in critically ill patients. *Crit Care Med* 2013; 41: 2373-2378.

19. van Vught LA, Klein Klouwenberg PM, Spitoni C, Scicluna BP, Wiewel MA, Horn J, Schultz MJ, Nurnberg P, Bonten MJ, Cremer OL, van der Poll T. Incidence, Risk Factors, and Attributable Mortality of Secondary Infections in the Intensive Care Unit After Admission for Sepsis. *Jama* 2016; 315: 1469-1479.

20. Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, Nurnberg P, Schultz MJ, Horn J, Cremer OL, Bonten MJ, Hinds CJ, Wong HR, Knight JC, van der Poll T. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *Lancet Respir Med* 2017; 5: 816-826.

21. Bos LD, Schouten LR, van Vught LA, Wiewel MA, Ong DSY, Cremer O, Artigas A, Martin-Loeches I, Hoogendijk AJ, van der Poll T, Horn J, Juffermans N, Calfee CS, Schultz MJ. Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis. *Thorax* 2017; 72: 876-883.

22. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34: 1297-1310.

23. Jones AE, Trzeciak S, Kline JA. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med* 2009; 37: 1649-1654.

24. Mason KL, Erb Downward JR, Mason KD, Falkowski NR, Eaton KA, Kao JY, Young VB, Huffnagle GB. Candida albicans and bacterial microbiota interactions in the cecum during recolonization following broad-spectrum antibiotic therapy. *Infect Immun* 2012; 80: 3371-3380.

25. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 2011; 108 Suppl 1: 4516-4522.

26. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013; 79: 5112-5120.

27. Oksanen JF, Blanchet G, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. vegan: Community Ecology Package. R package version 2.0-4.; 2012.

28. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

29. Wang Y, Naumann U, Wright ST, Warton DI. mvabund- an R package for model-based analysis of multivariate abundance data. *Methods Ecol Evol* 2012; 3: 471-474.

30. Schloss PD. MiSeq SOP - mothur. 2015. Available from: http://www.mothur.org/wiki/MiSeq_SOP.

31. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009; 75: 7537-7541.

32. Liaw A, Wiener M. Classification and Regression by randomForest. *R Sews* 2002; 2: 18-22.

33. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; 12: 87.

34. Dickson RP, Erb-Downward JR, Prescott HC, Martinez FJ, Curtis JL, Lama VN, Huffnagle GB. Cell-associated bacteria in the human lung microbiome. *Microbiome* 2014; 2: 28.

35. Schuijt TJ, Lankelma JM, Scicluna BP, de Sousa e Melo F, Roelofs JJ, de Boer JD, Hoogendijk AJ, de Beer R, de Vos A, Belzer C, de Vos WM, van der Poll T, Wiersinga WJ. The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut* 2016; 65: 575-583.

36. Emonet S, Lazarevic V, Leemann Refondini C, Gaia N, Leo S, Girard M, Nocquet Boyer V, Wozniak H, Despres L, Renzi G, Mostaguir K, Dupuis Lozeron E, Schrenzel J, Pugin J.

Identification of respiratory microbiota markers in ventilator-associated pneumonia. *Intensive Care Med* 2019.

37. Dickson RP, Erb-Downward JR, Huffnagle GB. Towards an ecology of the lung: new conceptual models of pulmonary microbiology and pneumonia pathogenesis. *Lancet Respir Med* 2014; 2: 238-246.

38. Schurink CAM, Nieuwenhoven CAV, Jacobs JA, Rozenberg-Arska M, Joore HCA, Buskens E, Hoepelman AIM, Bonten MJM. Clinical pulmonary infection score for ventilator-associated pneumonia: accuracy and inter-observer variability. *Intensive Care Med* 2004; 30: 217-224.

Online Data Supplement for “Lung microbiota predict clinical outcomes in critically ill patients”

Authors: Robert P. Dickson[1,2,3], Marcus J. Schultz[4,5], Tom van der Poll[6], Laura R. Schouten[4], Nicole R. Falkowski[1], Jenna E. Luth[1], Michael W. Sjoding[1,3,7], Christopher A. Brown[1], Rishi Chanderraj[1,8], Gary B. Huffnagle[1,2], Lieuwe D. J. Bos[4,9]

Affiliations: 1. Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan; 2. Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan; 3. Michigan Center for Integrative Research in Critical Care; Ann Arbor, Michigan; 4. Intensive Care, Amsterdam University Medical Centers, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands. 5. Mahidol University, Bangkok, Thailand: Mahidol-Oxford Tropical Medicine Research Unit (MORU). 6. Centre for experimental and molecular medicine (CEMM), Amsterdam University Medical Center, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands. 7. Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan; 8. Division of Infectious Diseases, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan; 9. Department of Respiratory Medicine, Amsterdam University Medical Centers, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands.

Supplemental Materials and Methods

Specimen processing

Cells were separated via centrifugation (15g for 15 min at 4 °C) and the cell-free supernatant was frozen at −80 °C for subsequent assays. Cell-free supernatants were subsequently centrifuged (22,500g for 30 min), and the resulting pellet was used for DNA isolation. Acellular mini-BAL pellets resuspended in 360µl ATL buffer (Qiagen DNeasy Blood & Tissue kit). Sterile laboratory water and AE buffer used in DNA isolation were collected and analyzed as potential sources of contamination, as were extraction controls (empty isolation tubes) and blank sequencing wells.

Bacterial DNA isolation

Genomic DNA was extracted from mini-BAL pellets (Qiagen DNeasy Blood & Tissue kit, Qiagen, Hilden, Germany) using a modified protocol previously demonstrated to isolate bacterial DNA(1). Sterile laboratory water and AE buffer used in DNA isolation were collected and analyzed as potential sources of contamination. Specimens were processed in a randomized order to minimize the risk of false pattern formation due to reagent contamination(2).

16s rRNA gene sequencing

The V4 region of the 16s rRNA gene was amplified using published primers(3) and the dual-indexing sequencing strategy developed by the laboratory of Patrick D. Schloss(4). Sequencing was performed using the Illumina MiSeq platform (San Diego, CA), using a MiSeq Reagent Kit V2 (500 cycles), according to the manufacturer's instructions with modifications found in the Schloss SOP(5). Accuprime High Fidelity Taq was used in place of Accuprime Pfx SuperMix. Primary PCR cycling conditions were 95°C for two minutes, followed by 20 cycles of touchdown PCR (95°C 20 seconds, 60°C 20 seconds and decreasing 0.3 degrees each cycle, 72°C 5 minutes), then 20 cycles

of standard PCR (95°C for 20 seconds, 55°C for 15 seconds, and 72°C for 5 minutes), and finished with 72°C for 10 minutes.

Bacterial DNA quantification

Bacterial DNA was quantified using a QX200 Droplet Digital PCR System (BioRad, Hercules, CA). Primers and cycling conditions were performed according to a previously published protocol(6). Specifically, primers were 5'- GCAGGCCTAACACATGCAAGTC-3' (63F) and 5'-CTGCTGCCTCCCGTAGGAGT-3' (355R). The cycling protocol was 1 cycle at 95°C for 5 minutes, 40 cycles at 95°C for 15 seconds and 60°C for 1 minute, 1 cycle at 4°C for 5 minutes, and 1 cycle at 90°C for 5 minutes all at a ramp rate of 2°C/second. The BioRad C1000 Touch Thermal Cycler was used for PCR cycling. Droplets were quantified using the Bio-Rad Quantisoft software. Two replicates were used per sample. No-template control specimens were used and were run alongside mini-BAL specimens.

Statistical analysis

Sequence data were processed and analyzed using the software mothur v.1.39.5 according to the Standard Operating Procedure for MiSeq sequence data using a minimum sequence length of 250 basepairs(5, 7). For each experiment and sequencing run, a shared community file and a phylotyped (genus-level grouping) file were generated using operational taxonomic units (OTUs) binned at 97% identity generated using the dist.seqs, cluster, make.shared and classify.otu commands in mothur. OTU numbers were arbitrarily assigned in the binning process and are referred to throughout the manuscript in association with their most specified level of taxonomy. Classification of OTUs was carried out using the mothur implementation of the Ribosomal Database Project (RDP) Classifier and the RDP taxonomy training set 14 (Trainset14_032015.rdp), available on the mothur website. Sequences are available via the NCBI Sequence Read Archive (accession

number PRJNA553560. OTU, taxonomy, and metadata tables are available at https://github.com/dicksonlunglab/MARS_lung_microbiome.

We performed microbial ecology analysis using the *vegan* package 2.4-1 and *mvabund* in *R*(8-10). For relative abundance and ordination analysis, samples were normalized to the percent of total reads and we restricted analysis to OTUs that were present at greater than 1% of the sample population. All OTUs were included in diversity analysis. Direct community similarity comparisons were performed using the Bray-Curtis similarity index. We performed ordinations using Principal Component Analysis on Hellinger-transformed normalized OTU tables generated using Euclidean distances(11). We determined significance of differences in community composition using PERMANOVA (*adonis*) with 10,000 permutations using Euclidean distances. We performed all analyses in *R* and GraphPad Prism 6. We compared means via Student's T test and ANOVA with Holm-Sidak's multiple comparisons test as appropriate.

Adequacy of sequencing and exclusion of specimens

Bacterial community analysis (using 16S rRNA gene amplicon sequencing) was performed on all specimens. We obtained 1,690,680 16S rRNA gene copies (18,578 ± 9,139 reads per specimen). Three specimens had inadequate amplification (<1000 16S rRNA gene copies) and were excluded from sequencing analysis (though included in ddPCR analysis). No bacterial taxa were excluded from analysis.

References:

1. Mason KL, Erb Downward JR, Mason KD, Falkowski NR, Eaton KA, Kao JY, Young VB, Huffnagle GB. Candida albicans and bacterial microbiota interactions in the cecum during recolonization following broad-spectrum antibiotic therapy. *Infect Immun* 2012; 80: 3371-3380.
2. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; 12: 87.
3. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 2011; 108 Suppl 1: 4516-4522.
4. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013; 79: 5112-5120.
5. Schloss PD. MiSeq SOP - mothur. 2015. Available from: http://www.mothur.org/wiki/MiSeq_SOP.
6. Sze MA, Abbasi M, Hogg JC, Sin DD. A comparison between droplet digital and quantitative PCR in the analysis of bacterial 16S load in lung tissue samples from control and COPD GOLD 2. *PLoS One* 2014; 9: e110351.
7. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009; 75: 7537-7541.
8. Oksanen JF, Blanchet G, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. vegan: Community Ecology Package. R package version 2.0-4.; 2012.
9. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
10. Wang Y, Naumann U, Wright ST, Warton DI. mvabund- an R package for model-based analysis of multivariate abundance data. *Methods Ecol Evol* 2012; 3: 471-474.

11. Legendre P, Gallagher ED. Ecologically meaningful transformations for ordination of species data. *Oecologia*; 129: 271-280.

Title: Lung microbiota predict clinical outcomes in critically ill patients

Authors: Robert P. Dickson[1,2,3], Marcus J. Schultz[4,5], Tom van der Poll[6], Laura R. Schouten[4], Nicole R. Falkowski[1], Jenna E. Luth[1], Michael W. Sjoding[1,3,7], Christopher A. Brown[1], Rishi Chanderraj[1,8], Gary B. Huffnagle[1,2], Lieuwe D. J. Bos[4,9]

Affiliations: 1. Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan; 2. Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan; 3. Michigan Center for Integrative Research in Critical Care; Ann Arbor, Michigan; 4. Intensive Care, Amsterdam University Medical Centers, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands. 5. Mahidol University, Bangkok, Thailand: Mahidol-Oxford Tropical Medicine Research Unit (MORU). 6. Centre for experimental and molecular medicine (CEMM), Amsterdam University Medical Center, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands. 7. Center for Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan; 8. Division of Infectious Diseases, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan; 9. Department of Respiratory Medicine, Amsterdam University Medical Centers, location AMC, University of Amsterdam, Infection and Immunity, Amsterdam, The Netherlands.

Collaborators: Published on behalf of the BASIC Consortium:

Amsterdam University Medical Centers, location AMC, Amsterdam, The Netherlands: de Beer FM, Bos LD, Claushuis TA, Glas GJ, Horn J, Hoogendijk AJ, van Hooijdonk RT, Huson MA, de Jong MD, Juffermans NP, Lagrand WA, van der Poll T, Scicluna B, Schouten LR, Schultz MJ, van der Sluijs KF, Straat M, van Vught LA, Wieske L, Wiewel MA, Witteveen E.

Corresponding Author Information:

Robert P. Dickson, MD
Pulmonary & Critical Care Medicine
University of Michigan Health System
6220 MSRB III / SPC 5642
1150 W. Medical Center Dr.
Ann Arbor, MI 48109-5642
rodickso@med.umich.edu
(734)277-1487 (mobile)
(734)764-2655 (fax)
(734)936-9368 (lab)
(734)936-5010 (administrator)

Author Contributions: Conception and design: RPD, LDJB. Acquisition of data: MJS, LRS, TVP, LDJB. Processing of specimens and generation of data: NRF, JEL. Analysis and interpretation of data: RPD, MWS, CAB, RC, LDJB. Drafting or revising of manuscript: RPD, MJS, LRS, TVP, NRF, JEL, MWS, CAB, RC, GBH, LDJB. Final approval of manuscript: RPD, MJS, LRS, TVP, NRF, JEL, MWS, CAB, RC, GBH, LDJB.

Funding: Sample and clinical data collection was supported by the Center of Translational Medicine via the MARS consortium. Additional support came from the Academic Medical Center. The authors were supported by the National Institutes for Health (K23HL130641 [RPD], R21AI137669 [RPD], R01HL144599 [RPD]. Additional support was provided by the University of Michigan Center for Integrative Research in Critical Care [RPD].

Running Head: Lung microbiota predict ICU outcomes

Descriptor: 10.9 Pathogen/Host Cell Interactions

Word Count: 3495

At a Glance Commentary

Scientific Knowledge on the Subject

Recent studies have revealed that the lung microbiota of critically ill patients are profoundly altered and are correlated with alveolar and systemic inflammation. Altered lung microbiota may propel and perpetuate alveolar inflammation and injury among critically ill patients. Yet to date, no study has determined whether altered lung microbiota predict disease outcomes in this population.

What This Study Adds to the Field

We here show that among mechanically ventilated critically ill patients, variation in lung microbiota at admission predicts ICU outcomes. Two key features of the lung microbiome, bacterial burden and community composition, predict ventilator-free days. Specifically, increased lung bacterial DNA burden and enrichment of the lung microbiome with gut-associated bacterial taxa (e.g. *Lachnospiraceae* and *Enterobacteriaceae* families) were predictive both of poor ICU outcomes and the clinical diagnosis of ARDS. This correlation between gut-associated bacteria and ARDS validates prior findings, and supports the hypothesis that translocation of gut bacteria to the lungs contributes to the pathogenesis of lung injury. Our results confirm the clinical significance of lung microbiota in critically ill patients. The lung microbiome is an important and underappreciated source of clinical heterogeneity among the critically ill, and represents a novel therapeutic target for the prevention and treatment of lung injury.

Abstract

Rationale: Recent studies have revealed that in critically ill patients, lung microbiota are altered and correlate with alveolar inflammation. The clinical significance of altered lung bacteria in critical illness is unknown.

Objectives: To determine if clinical outcomes of critically ill patients are predicted by features of the lung microbiome at the time of admission.

Methods: We performed a prospective observational cohort study in an intensive care unit (ICU) at a university hospital. Lung microbiota were quantified and characterized using droplet digital PCR and bacterial 16S rRNA gene sequencing. Primary predictors were the bacterial burden, community diversity, and community composition of lung microbiota. The primary outcome was ventilator-free days, determined at 28 days post admission.

Measurements and Main Results: Lungs of 91 critically ill patients were sampled using miniature-bronchoalveolar lavage within 24 hours of ICU admission. Patients with increased bacterial lung bacterial burden had fewer ventilator-free days (HR 0.43, CI 0.21-0.88), which remained significant when controlled for pneumonia and severity of illness. The community composition of lung bacteria predicted ventilator-free days (P=0.003), driven by the presence of gut-associated bacteria (e.g. *Lachnospiraceae* and *Enterobacteriaceae* spp.). Detection of gut-associated bacteria was also associated with the presence of the acute respiratory distress syndrome.

Conclusions: Key features of the lung microbiome (bacterial burden, enrichment with gut-associated bacteria) predict outcomes in critically ill patients. The lung microbiome is an understudied source of clinical variation in critical illness, and represents a novel therapeutic target for the prevention and treatment of acute respiratory failure.

Body

Introduction

In the past decade, advances in culture-independent microbiology have revealed that the lungs, previously considered sterile, harbor complex and dynamic communities of bacteria~~[1]~~(1). Lung microbiota are detectable in health~~[2-4]~~(2-4), altered in disease(5, 6)~~[5, 6]~~, and correlate with variation in airway and alveolar immunity~~[2, 4, 7]~~(2, 4, 7). In numerous chronic respiratory diseases, key features of the lung microbiome are predictive of disease outcomes. The burden of lung bacteria (measured by quantification of bacterial DNA) predicts mortality and disease progression in stable patients with idiopathic pulmonary fibrosis(8, 9) ~~[8, 9]~~and responsiveness to inhaled antibiotics in patients with bronchiectasis(10). The diversity of sputum microbiota predicts mortality in patients with chronic obstructive pulmonary disease(11)~~[10]~~, and the community composition of respiratory microbiota predicts exacerbations in bronchiectasis(12) ~~[11]~~ and respiratory infections in infants(13)~~[12]~~.

The lung microbiota of critically ill patients are profoundly altered compared to those of healthy subjects(7, 14-16), and correlate with alveolar and systemic inflammation(7, 15). Specifically, among patients with the acute respiratory distress syndrome (ARDS), the lung microbiome is enriched with gut-associated bacteria(7), and early enrichment of the lung microbiome with gut-associated bacteria (e.g. *Enterobacteriaceae* spp.) is associated with subsequent development of ARDS(15). Altered lung microbiota may propel and perpetuate alveolar inflammation and injury among critically ill patients, but to date no study has determined whether altered lung microbiota predict disease outcomes in this population.

To determine if lung microbiota at ICU admission predict clinical outcomes in critically ill patients, we performed a prospective observational cohort study on critically ill patients receiving mechanical ventilation. The primary outcome was ventilator-free days, adjudicated at

28 days following enrollment. We hypothesized that key features of the lung microbiome (bacterial burden, diversity, and community composition) would predict ICU outcomes, even when controlled for the presence of clinically-appreciated pneumonia.

Methods

Study design

This study was a secondary analysis of specimens collected from patients in the BASIC study (Biomarker Analysis in Septic ICu patients). This study was incorporated in the Molecular Diagnosis and Risk Stratification of Sepsis (MARS) project (NCT01905033) (17-20). The present study was conducted in the ICU of the Academic Medical Center and was approved by the institutional Medical Ethics committee; written informed consent was obtained from the patient representative prior to collection of airway samples via miniature bronchoalveolar lavage (IRB no. NL34294.018.10). Analysis was restricted to baseline specimens and data collected within 24 hours of admission.

Study population

All patients older than 18 years admitted to the ICU with an expected length of stay longer than 24 hours were included in the MARS project. The BASIC study comprised a subset of patients included in the MARS study at the Amsterdam ICU with at least two “systemic inflammatory response syndrome” criteria, who received no antibiotics in the days preceding ICU admission. The current analysis is limited to consecutive patients who were included between September 12, 2011, and November 7, 2013, receiving invasive mechanical ventilation and with informed consent for distal airway sampling. Adjudication of infection was assessed retrospectively using a four-point scale (ascending from none, possible, probable, to definite) using the Centers for Disease Control and Prevention and International Sepsis Forum consensus definitions as previously described(18). ARDS was scored on a daily basis by a team of well-trained clinical

researchers according to the American-European consensus criteria. After the publication of the Berlin definition, all cases were re-evaluated scored according to the new definition, as described previously(21). For the purposes of ARDS vs non-ARDS comparisons, we used adjudication at 24 hours following ICU admission. Severity of illness was quantified using the validated APACHE IV (Acute Physiology and Chronic Health Evaluation IV) (22) and SOFA (Sequential Organ Failure Assessment)(23) models.

Specimen collection and processing

Miniature bronchoalveolar lavage specimens were collected using standard clinical protocol. In short, a 50 cm, 14 Fr tracheal suction catheter was introduced through the orotracheal tube and inserted until significant resistance was encountered. The catheter was then pulled back 1cm and 20 mL 0.9% saline was injected in 10 seconds and immediately aspirated, after which the catheter was removed. Specimens were stored on ice from the time of specimen collection until processing. DNA was extracted, amplified, and sequenced according to previously published protocols(24-26). Sequencing was performed using the Illumina MiSeq platform (San Diego, CA). Bacterial DNA was quantified using a QX200 Droplet Digital PCR System (BioRad, Hercules, CA). Additional details are provided in the online data supplement.

Statistical analysis

As detailed in the online data supplement, we performed microbial ecology analysis using the *vegan* package 2.4-1 and *mvabund* in *R*(27-29) following sequence processing with *mothur*(30, 31). We pre-specified that key features of the microbiome (predictors) would be: 1) bacterial DNA burden (as quantified using droplet digital PCR), 2) bacterial community diversity (as calculated using the Shannon Diversity Index), and 3) community composition. We determined significance in community composition (e.g. mini-BAL specimens vs negative sequencing controls, ARDS vs non-ARDS mini-BAL specimens) using *mvabund* (model-based approach to

analysis of multivariate abundance data). To identify community members driving differences in community composition, we used 1) biplot analysis, 2) rank abundance analysis, and 3) a random forest ensemble learning approach (*randomForest* package in R, version 4.6-14(32)). For random forest, we determined variable importance using 100 forests. The importance parameter was set to TRUE to enable retrieval of importance metrics: mean decrease in accuracy for classification (ARDS vs. non-ARDS) or IncMSE for regression (ventilator-free days). Default settings were utilized for all other parameters. Following model creation, the unscaled feature importance metric was extracted from each forest, assembled into a dataframe, ordered by highest feature importance, and displayed in boxplots of the most important features across the 100 forests. Our primary index of feature importance was Mean Decrease in Accuracy, which ranks predictors by the relative loss in accuracy the occurs when they are removed from the predictive model. ~~For random forest, we determined variable importance using 100 forests and default settings for all parameters.~~ We compared means via Student's T test (when normally distributed), the Mann-Whitney U test (when non-Gaussian), and ANOVA with Holm-Sidak's multiple comparisons test as appropriate. Time-to-event analysis was performed using univariate and multivariate Cox proportional hazard models using ventilator-free days (adjudicated 28 days following enrollment) as a primary outcome; multivariate analysis adjusted for age, sex, severity of illness (APACHE IV), diagnosis of ARDS, and the presence of clinically-suspected pneumonia as determined both by the primary clinical service and via post-hoc CDC adjudication criteria. The primary outcome was the proportional hazard ratio for being alive and liberated from mechanical ventilation, as adjudicated 28 days following admission.

Role of the funding source

The funding agencies had no role in the design, conduct, and analysis of the study or in the decision to submit the manuscript for publication.

1. Sibila O, Laserna E, Shoemark A, Keir HR, Finch S, Rodrigo-Troyano A, Perea L, Lonergan M, Goeminne PC, Chalmers JD. Airway Bacterial Load and Inhaled Antibiotic Response in Bronchiectasis. *Am J Respir Crit Care Med* 2019; 200: 33-41.

Results

Study cohort

We obtained mini-BAL specimens from 91 critically ill patients within 24 hours of their ICU admission. The consort diagram is shown in the Appendix Figure E1. Patient demographics and clinical characteristics are reported in **Table 1**.

Table 1. Demographics and Clinical Characteristics of Study Cohort

Characteristic	Study cohort (*n* = 91)
Mean age (SD), *y*	60.7 (15.4)
Male	55 (60)
Admission type	
Medical	67 (74)
Surgical (emergency)	20 (22)
Surgical (elective)	4 (4)
Severity of illness	
Mean SOFA (SD)	7.2 (4.1)
Mean APACHE IV (SD)	82.6 (28.5)
Lung injury	
ARDS at admission	17 (19)
Mean PaO_2:FiO_2 (SD)	262.0 (104.7)
Comorbidities	
Diabetes mellitus	13 (14)
Malignancy	11 (12)
COPD	5 (5)
Immune deficiency	4 (4)
ICU outcomes	
Mean ventilator-free days (SD)	18.5 (10.5)
Mean ICU length of stay, days (SD)	5.6 (4.6)
30-day mortality	27 (30)

SOFA = Sequential Organ Failure Score. APACHE IV = Acute Physiology and Chronic Health Score IV. ARDS = Acute Respiratory Distress Syndrome. COPD = Chronic Obstructive Pulmonary Syndrome. Values are numbers (percentages) unless otherwise indicated.

~~**Table 1.** Demographics and Clinical Characteristics of Study Cohort~~

~~Characteristic~~	~~Study cohort (*n* = 91)~~
~~Mean age (SD), *y*~~	~~60.7 (15.4)~~

Male	55 (60)
Admission type	
Medical	67 (74)
Surgical (emergency)	20 (22)
Surgical (elective)	4 (4)
Severity of illness	
Mean SOFA (SD)	7.2 (4.1)
Mean APACHE IV (SD)	82.6 (28.5)
Lung injury	
ARDS at admission	17 (19)
Mean PaO_2:FiO_2 (SD)	262.0 (104.7)
Comorbidities	
COPD	5 (5)
Immune deficiency	4 (4)
ICU outcomes	
Mean ventilator-free days (SD)	18.5 (10.5)
Mean ICU length of stay, days (SD)	5.6 (4.6)
30-day mortality	27 (30)

SOFA = Sequential Organ Failure Score. APACHE IV = Acute Physiology and Chronic Health Score IV. ARDS = Acute Respiratory Distress Syndrome. COPD = Chronic Obstructive Pulmonary Syndrome. Values are numbers (percentages) unless otherwise indicated.

The distribution of admission diagnoses is reported in **Appendix Table E1**. Bacterial quantification and 16S rRNA gene sequencing was performed on all specimens. Details regarding adequacy of sequencing and exclusion of specimens are provided in the online supplement.

The microbiota of lung specimens from critically ill patients are distinct from those of background sequencing controls

Low-biomass microbiome studies are vulnerable to contamination due to bacterial DNA present in reagents used in DNA extraction and library preparation(33). Our study used low-volume specimens (mini-BAL) from a low-microbial-biomass body site (the lower respiratory tract), and had their bacterial burden further decreased via a centrifugation step to remove eukaryotic cells(34). Given these concerns, we first asked whether a bacterial signal could be detected in

these mini-BAL specimens that was distinct from that of negative controls. We accomplished this by comparing the bacterial DNA in mini-BAL specimens with no-template controls (N = 25), AE buffer specimens (N = 8), sterile water used in DNA extraction (N = 4), extraction control specimens (N = 9), and blank sequencing wells (N = 6).

As shown in **Supplemental Figure ~~1~~2**, we found clear evidence of distinct bacterial signal in mini-BAL specimens despite their low microbial biomass. Using ultrasensitive quantification of the bacterial 16S gene (via droplet digital PCR), we found significantly more bacterial DNA in mini-BAL specimens than in no-template control specimens ($P < 0.001$, **Supplemental Figure ~~1A~~2A**). The median bacterial DNA burden was 21,246 bacterial gene copies per mL (mean 118,411 copies ± 707,438). We found a wide range of bacterial burden (6,329 - 6,713,947 copies/mL), ranging from comparable to no-template controls to 1000-fold greater than background. Using bacterial community analysis (16S rRNA gene amplicon sequencing), we confirmed that the identity of bacteria detected in mini-BAL specimens was distinct from that of negative control specimens ($P < 0.0001$, *mvabund*). Principal component analysis revealed distinct clustering of mini-BAL specimens apart from negative control specimens (**Supplemental Figure ~~1B~~2B**), though overlap did occur between some mini-BAL specimens and negative controls. Rank abundance analysis showed clear differences in relative abundance of taxa in negative controls and mini-BAL specimens (**Supplemental Figure ~~1C~~2C**). The dominant taxonomic group in negative controls specimens (OTU008:*Pelomonas*) comprised 25.5% of bacterial sequences in negative controls, but only 2.6% of sequences in mini-BAL specimens.

We thus concluded that while highly variable in their bacterial DNA burden and susceptibility to contamination, mini-BAL specimens contained a distinct bacterial signal from negative control specimens.

Lung microbiota of critically ill patients are altered in patients with ARDS and enriched with gut-associated bacteria (*Enterobacteriaceae* spp.)

We next compared the lung microbiota of critically ill patients with and without ARDS. Prior studies have demonstrated that the lung microbiota of patients with ARDS are altered and enriched with gut-associated bacteria. We compared lung bacterial communities in patients with and without physician-adjudicated ARDS. As shown in **Figure 1**, lung bacterial communities of patients with ARDS differed in the bacterial DNA burden and community composition compared to patients without ARDS.

We first compared ARDS and non-ARDS specimens in their bacterial DNA burden and community diversity. Though highly variable, the bacterial DNA burden in mini-BAL specimens was greater in patients with ARDS than without ARDS (P = 0.014, **Figure 1A**). ARDS specimens did not differ in bacterial community diversity, either measured via the Shannon Diversity Index (P = 0.13) or community richness (P = 0.83) (**Figure 1B**). With both comparisons (bacterial DNA burden and diversity), within-group variation far exceeded across-group differences.

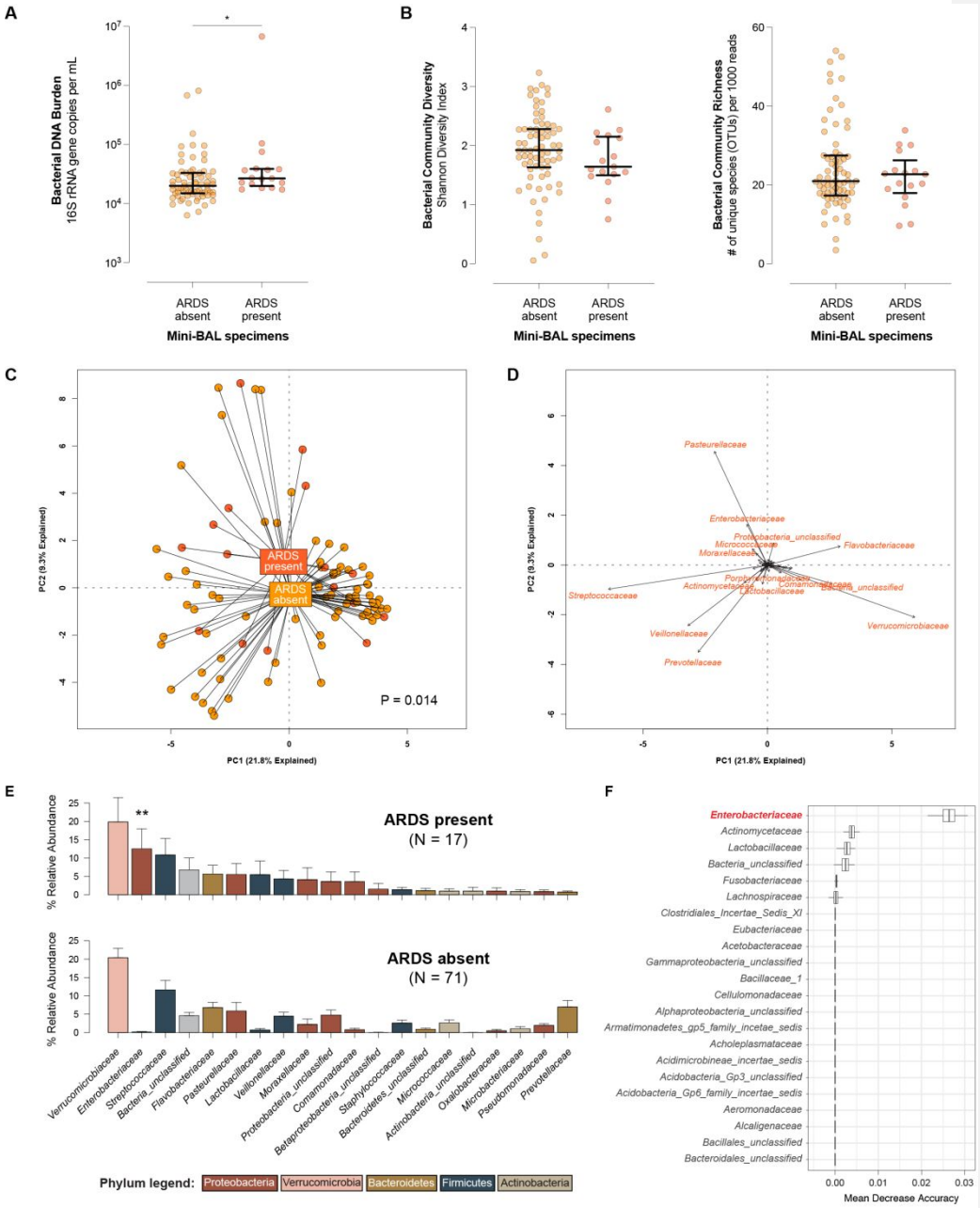


Figure 1. Lung microbiota are altered in patients with the Acute Respiratory Distress Syndrome (ARDS). Compared to patients without ARDS, patients with ARDS had increased lung bacterial burden (A) but no difference in community diversity (B). Principal component analysis of bacterial communities (C) revealed that the community composition of lung bacteria was distinct in specimens from ARDS patients (C), driven by members of the Pasteurellaceae and Enterobacteriaceae families (D). Rank abundance analysis (E) identified taxa enriched in ARDS specimens (e.g. Enterobacteriaceae), and Random Forest analysis (F) confirmed that the Enterobacteriaceae family was the most discriminating taxonomic group between ARDS and non-ARDS specimens. Hypothesis testing performed using Mann-Whitney U test (A), Student's t test (B) and mvabund (C, E). A, B: median and interquartile range.

We next compared the community composition of bacterial communities in ARDS and non-ARDS specimens using complementary approaches. We first visualized communities using principal component analysis (**Figure 1C**). While considerable taxonomic overlap was found across ARDS and non-ARDS specimens, there was a detectable separation of specimens according to ARDS status. This collective difference in community composition was confirmed statistically via *mvabund*, and was robust to taxonomic level of comparison (P = 0.014 at the OTU level of taxonomy, P = 0.013 at the family level, P = 0.003 at the phylum level). We next used biplot analysis to identify specific taxa responsible for this collective difference in community composition (**Figure 1D**). Whereas clusters of non-ARDS specimens were defined by bacterial taxa commonly detected in healthy lungs (*Streptococcaceae* spp., *Veillonellaceae* spp., and *Prevotellaceae* spp.) and taxa detected in negative sequencing control specimens (*Verrucomicrobiaceae* spp., *Flavobacteriaceae* spp.), ARDS specimens were more commonly characterized by *Pasteurellaceae* spp. and *Enterobacteriacaeae* spp..

We then used complementary techniques to identify ARDS-associated bacterial taxa. Using rank abundance visualization (**Figure 1E**), we compared the relative abundance of prominent taxa across ARDS and non-ARDS specimens. While many taxa were common to both group, the *Enterobacteriaceae* family was far more abundant ARDS specimens compared to non-ARDS specimens (12.5% of all bacterial sequences in ARDS specimens compared to 0.18% of all bacterial specimens in non-ARDS specimens). We used unbiased regression-based (*mvabund*) and ensemble-learning (*random forest*) approaches to identify ARDS-enriched taxa. *Mvabund*, which rigorously controls for multiple comparisons, identified the *Enterobacteriaceae* family as enriched in ARDS specimens (P = 0.002). *Random forest* clearly identified the *Enterobacteriaceae* family as the most important taxonomic feature discriminating ARDS from non-ARDS specimens (**Figure 1F**).

We next compared our ARDS-associated *Enterobacteriaceae* taxonomic group with that of an ARDS-associated ~~*Enterobacteriacae*~~*Enterobacteriaceae* taxon in a recently published study of mechanically ventilated trauma patients(15). We compared the most prominent *Enterobacteriaceae*-classified OTU in our data set (OTU0005, comprising 61.5% of all *Enterobacteriaceae*-classified sequences) with the ARDS-associated *Enterobacteriaceae* identified by Panzer et al. (OTU2119418). As shown in **Supplemental Figure ~~2A~~3A**, the representative sequence of our study's ARDS-associated *Enterobacteriaceae* OTU was 96% aligned with that of the ARDS-associated *Enterobacteriaceae* OTU identified by Panzer et al., differing in only 3 base pairs. We compared these ARDS-associated OTUs with the taxonomic classifications of closely-aligned sequences from the SILVA ribosomal RNA database. As shown in in **Supplemental Figure ~~2B~~3B**, both OTUs were exclusively identical to *Enterobacteriaceae*-classified taxa, including *Escherichia coli*, *Enterobacter* spp., and *Klebsiella pneumoniae*.

We thus concluded that the lung microbiota of patients with ARDS differ from those of critically ill patients without ARDS, driven by relative enrichment with gut-associated *Enterobacteriaceae* spp..

Lung microbiota are predictive of clinical outcomes in critically ill patients

We next asked if key features of the lung microbiome (*bacterial burden*, *diversity*, and *community composition*) predict clinical outcomes in critically ill patients. Our primary outcome was ventilator-free days measured at 28 days following admission.

We first asked if *bacterial burden* of mini-BAL specimens (quantified using ddPCR of the 16S rRNA gene) predicted ICU outcomes (**Table 2**). Using univariate analysis, we found that increased baseline lung bacterial DNA burden predicted fewer ventilator-free days, either when

analyzed continuously (hazard ratio 0.43, confidence interval 0.21 - 0.88, P = 0.022) or when comparing tertiles defined by total lung bacterial DNA burden. In other words, for each additional 10-fold increase in lung bacterial DNA, the Hazard Ratio for favorable outcome (liberation from mechanical ventilation) was 0.43. As shown in **Figure 2**, the tertile of patients with the highest baseline lung bacterial DNA burden were less likely to be extubated and alive at 7, 14, 21, and 28 days compared to patients with low bacterial DNA burden (HR 0.45, confidence interval 0.25 - 0.81, P = 0.008).

Table 2. Predictors of Ventilator-Free Days in Mechanically Ventilated Critically Ill Patients

Predictor	*Univariate*		*Multivariate*	
	Hazard ratio (CI)	***P* value**	**Hazard ratio (CI)**	***P* value**
Lung bacterial DNA burden (continuous)	0.43 (0.21 - 0.88)	0.022	0.40 (0.18 - 0.86)	0.019
Lung bacterial DNA burden: middle tertile*	0.87 (0.50 - 1.51)	0.62		
Lung bacterial DNA burden: highest tertile*	0.45 (0.25 - 0.81)	0.008		
Shannon diversity index†	1.27 (0.87 - 1.86)	0.21		
Age (years)	0.99 (0.98 - 1.01)	0.35	1.01 (0.99 - 1.03)	0.32
Gender (male)	1.26 (0.78 - 2.03)	0.35	0.90 (0.54 - 1.49)	0.68
SOFA	0.95 (0.90 - 1.01)	0.10		
APACHE IV	0.98 (0.98 - 0.99)	<0.001	0.98 (0.97 - 0.99)	<0.001
Suspected pneumonia	1.01 (0.60 - 1.70)	0.96	0.90 (0.53 - 1.55)	0.71
Pneumonia (post-hoc, CDC criteria)	0.48 (0.18 - 1.33)	0.16		
ARDS	0.51 (0.27 - 0.98)	0.044	0.61 (0.31 - 1.21)	0.16

CI = Confidence interval (95%). SOFA = Sequential Organ Failure Score. APACHE IV = Acute Physiology and Chronic Health Score IV. ARDS = Acute Respiratory Distress Syndrome.
*Versus lowest tertile.
†Shannon diversity index, per 1-unit increase

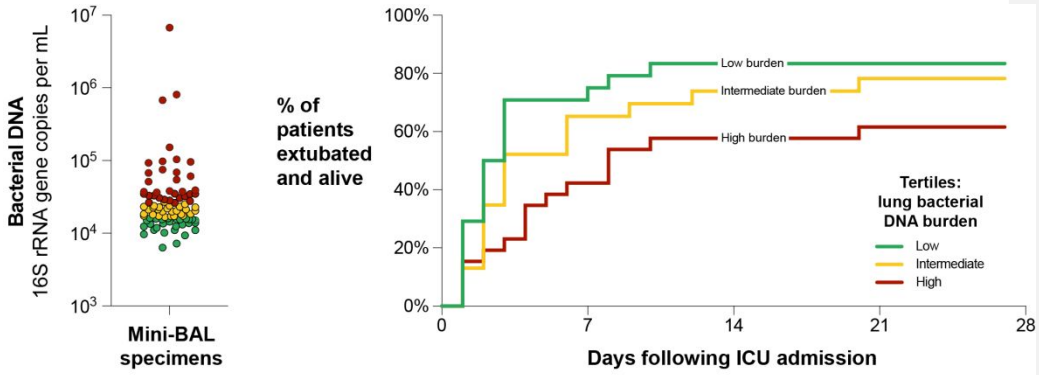


Figure 2. Lung microbiota predict 28-day outcomes in mechanically-ventilated critically ill patients. In critically ill patients receiving mechanical ventilation, the burden of bacterial DNA detected in mini-BAL specimens was predictive of total ventilator-free days. Patients with high lung burdens of bacterial DNA were less likely to be extubated and alive than patients with low bacterial DNA burden (P = 0.008). Hypothesis testing performed using univariate Cox proportional hazard modeling.

Pneumonia is common among mechanically ventilated patients, and a potential source of confounding in lung microbiome studies. In our study cohort, 27 patients (30%) had suspected pneumonia either at admission or within 48 hours; of these, 7 patients met CDC criteria for probable or confirmed pneumonia via post-hoc chart review. When we controlled our outcome analysis of lung bacterial DNA burden for the presence of clinically suspected pneumonia, it did not meaningfully change either the hazard ratio or significance of the model (HR = 0.43, P = 0.021). Similarly, controlling for post-hoc adjudicated pneumonia (using CDC criteria) also did not influence the predictive power of lung bacterial DNA burden (HR = 0.43, P = 0.019). We thus concluded that lung bacterial burden predicts poor outcomes in mechanically ventilated critically ill patients, even when controlled for the presence of suspected or confirmed pneumonia.

We then performed multivariate analysis to determine whether lung bacterial DNA burden is independently predictive of poor outcomes. The relationship between increased lung bacterial DNA burden and fewer ventilator-free days remained significant when controlled for age,

gender, severity of illness (APACHE IV score), the presence of suspected pneumonia, and the presence of ARDS (HR 0.40, confidence interval 0.18 - 0.86, P = 0.019). We thus concluded that lung bacterial DNA burden is an independent predictor of poor outcomes in critically ill patients.

We next asked if *bacterial diversity* of lung bacteria predicts ICU outcomes (**Figure 3A**). Bacterial diversity of lung bacteria (as measured by the Shannon Diversity Index) did not significantly predict ICU outcomes (P = 0.22). While the most favorable ICU outcomes were observed among patients with high baseline lung bacterial diversity, followed in stepwise manner by patients with intermediate and low diversity, this difference in tertiles was not statistically significant. Other indices of lung bacterial diversity (community richness, community dominance) were also not significantly predictive of ICU outcomes (ventilator-free days, P > 0.05 for all comparisons, **Appendix Table ~~E1~~E2**).

~~We then performed multivariate analysis to determine whether lung bacterial DNA burden is independently predictive of poor outcomes. The relationship between increased lung bacterial DNA burden and fewer ventilator-free days remained significant when controlled for age, gender, severity of illness (APACHE IV score), the presence of suspected pneumonia, and the presence of ARDS (HR 0.40, confidence interval 0.18 - 0.86, P = 0.019). We thus concluded that lung bacterial DNA burden is an independent predictor of poor outcomes in critically ill patients.~~

Finally, we asked if the *community composition* of lung bacteria is predictive of ICU outcomes. We compared patient ventilator-free days with lung bacterial community structure using *mvabund* (model-based approach to analysis of multivariate abundance data). The overall community composition of baseline lung microbiota was significantly predictive of patient

ventilator-free days (P = 0.003 at the OTU level of taxonomy, P = 0.004 at the family level). Using *random forest* to identify taxa associated with poor outcomes, we identified the gut-associated *Lachnospiraceae* and *Enterobacteriaceae* families as the taxa most strongly predictive of fewer ventilator-free days (**Figure 3B**). We then tested the hypotheses that detection of these taxa predicts poor outcomes. As shown in **Figure 3B**, detection of the *Lachnospiracaeae* family was significantly predictive of worse ICU outcomes (P = 0.020). The relationship between *Enterobacteriaceae* detection and ventilator-free days was not significant (P = 0.12).

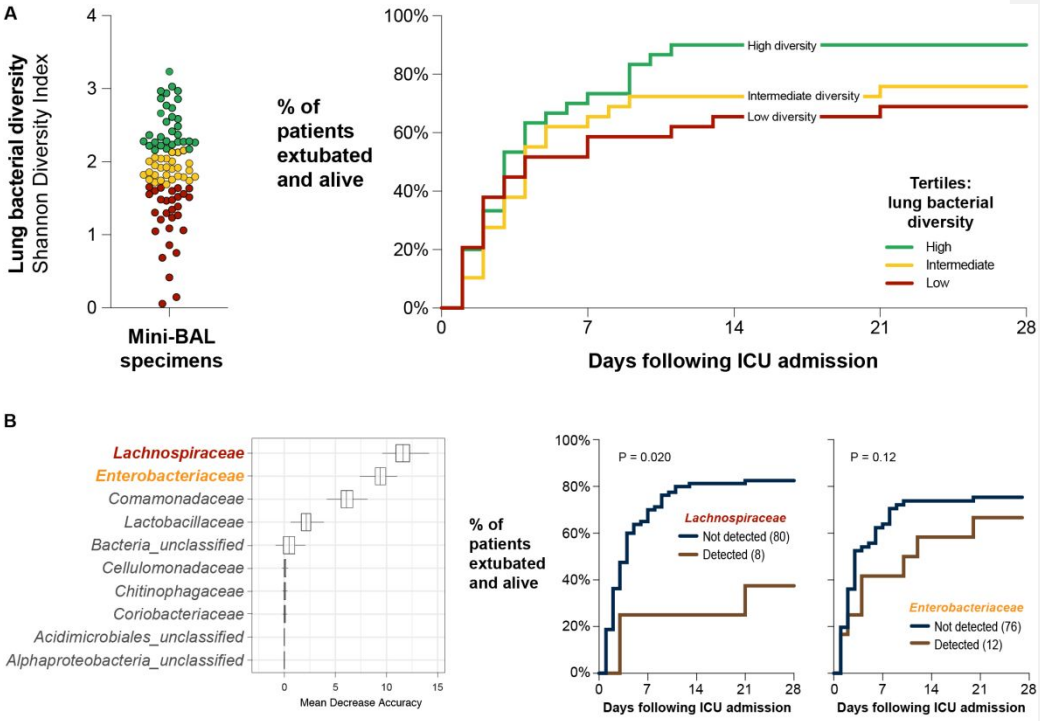


Figure 3. Lung microbiota and 28-day outcomes in mechanically-ventilated critically ill patients. Community diversity of lung bacteria was highly variable among patients, and did not significantly predict ventilator-free days (A). Community composition of lung bacteria was significantly predictive of ventilator-free days (P = 0.003, mvabund). Random forest identified the gut-associated *Lachnospiracea* and *Enterobacteriaceae* families as the strongest predictors of ventilator-free days (B). Hypothesis testing performed using Cox proportional hazard model.

We thus concluded that in the lung microbiota of critically ill patients, poor ICU outcomes are predicted both by increased bacterial burden and by community composition (specifically, enrichment with gut-associated taxa).

Discussion

The core finding of this study is that among mechanically ventilated critically ill patients, variation in lung microbiota at admission predicts ICU outcomes. Two key features of the lung microbiome, bacterial burden and community composition, predicted ventilator-free days. Specifically, increased lung bacterial DNA burden and enrichment of the lung microbiome with

gut-associated bacterial taxa (e.g. *Lachnospiraceae* and *Enterobacteriaceae* families) were predictive of poor ICU outcomes and the clinical diagnosis of ARDS.

Our study is the first to demonstrate that variation in lung microbiota is predictive of clinical outcomes in critically ill patients. This key finding is compatible with broader (non-ICU) lung microbiome studies, which have found that lung microbiota are predictive of disease outcomes in idiopathic pulmonary fibrosis(8, 9), chronic obstructive pulmonary disease(11), bronchiectasis(10, 12), and infants susceptible to respiratory infections(13). While these findings, robust across disease states, confirm that the lung microbiome is a risk factor for disease progression, a crucial and unanswered question is whether lung microbiota are a *modifiable* risk factor. Animal studies using germ-free and antibiotic-exposed mice suggest that manipulation of the microbiome does influence host susceptibility to lung inflammation, injury, and mortality(9, 35). Yet the precise role of gut and lung microbiota in the pathogenesis of acute and chronic lung injury has yet to be elucidated. Future studies should interrogate whether the microbiome's role in lung disease is more attributable to remote (gut-lung) or local (lung-lung) host-microbiome interactions(2).

Our findings both validate several recent studies and provide new insight into the importance of the lung microbiome in critical illness. We have previously reported that the lung microbiome is enriched with gut-associated bacteria in ARDS and correlated with the severity of systemic and alveolar inflammation(7). In a subsequent study of mechanically ventilated trauma patients, the presence of gut-associated bacteria in endotracheal aspirates (*Enterobacteriaceae* spp.) was associated with ARDS onset(15). In our current study, we found that the lung microbiota of patients with ARDS was distinct from patients without ARDS, again driven by the presence of gut-associated bacteria (*Enterobacteriaceae* spp.). Indeed, the bacterial taxon most strongly correlated with ARDS status in our study (OTU0005:*Enterobacteriaceae*) was nearly identical to

that of the ARDS-associated bacterial taxon found by Panzer and colleagues(15). These multiple findings, now robust across cohorts, sequencing platforms, laboratories, and continents, all provide indirect support for the hypothesis that gut-lung translocation of bacteria contributes to the pathogenesis of lung injury in critically ill patients.

Importantly, our core findings remained significant when controlled for the clinical suspicion or post-hoc adjudication of pneumonia. While this may seem paradoxical (increased lung bacterial burden predicts poor outcomes, and pneumonia is a condition of high lung bacterial burden), recent culture-independent studies have revealed both the complexity of lung bacterial communities in mechanically ventilated patients(14, 16, 36) and the inadequacy of our conventional understanding of pneumonia(37). The lack of concordance between our molecular characterization of lung bacteria and clinical assessment of pneumonia likely reflects several key issues in the microbiology of injured lungs: 1) clinical adjudication of pneumonia, especially in mechanically ventilated patients, is imprecise and unreliable(38), 2) a dichotomous adjudication of pneumonia is too simplistic and reductionistic to meaningfully describe the complex ecologic spectrum of respiratory microbiota, and 3) the lung microbiome may play a role in the pathogenesis of disease processes not classically considered infectious (e.g. perpetuating inflammation and injury in ARDS). Our results highlight the need for improved molecular diagnostics to provide clinicians with a more accurate and comprehensive assessment of lung microbiota, as well as a more refined ecologic understanding of respiratory infections in critically ill patients.

Our study has several limitations that should prompt further validation and study. While we detected a distinct bacterial signal in our specimens, the bacterial biomass in these cell-free mini-BAL specimens was low, and in many specimens overlapped with background “sequencing noise.” Future studies, using larger volumes of whole BAL, ~~will likely find even~~ may find stronger

bacterial signal. Our mini-BAL sampling approach was non-directional; thus the anatomic site of sampling was not standardized across patients. While our findings remained significant when controlled for important clinical confounders, we could not control for all potential exposures (e.g. ICU antibiotic exposure or pre-ICU medications), and residual confounding is likely. Finally, while our findings provide indirect support for the hypothesis of gut-lung translocation contributing to lung injury in critically ill patients, our lack of paired gut specimens precludes our determining whether gut-associated taxa (e.g. *Enterobacteriaceae* and *Lachnospiracaeae* spp.) were derived from the lower gastrointestinal tract or via another route (e.g. aspiration of altered pharyngeal microbiota). Future prospective studies of critically ill patients, in addition to sampling the lower respiratory tract, should collect time-matched specimens from the lower and upper gastrointestinal tract.

In conclusion, in this prospective observational cohort study of mechanically ventilated critically ill patients, variation in baseline lung microbiota predicted ICU outcomes. Increased lung bacterial burden and lung enrichment with gut-associated bacteria were predictive of worse outcomes. The lung microbiome is an important and understudied source of variation among critically ill patients, and may represent a novel therapeutic target for the prevention and treatment of lung injury.

406 **Data availability**

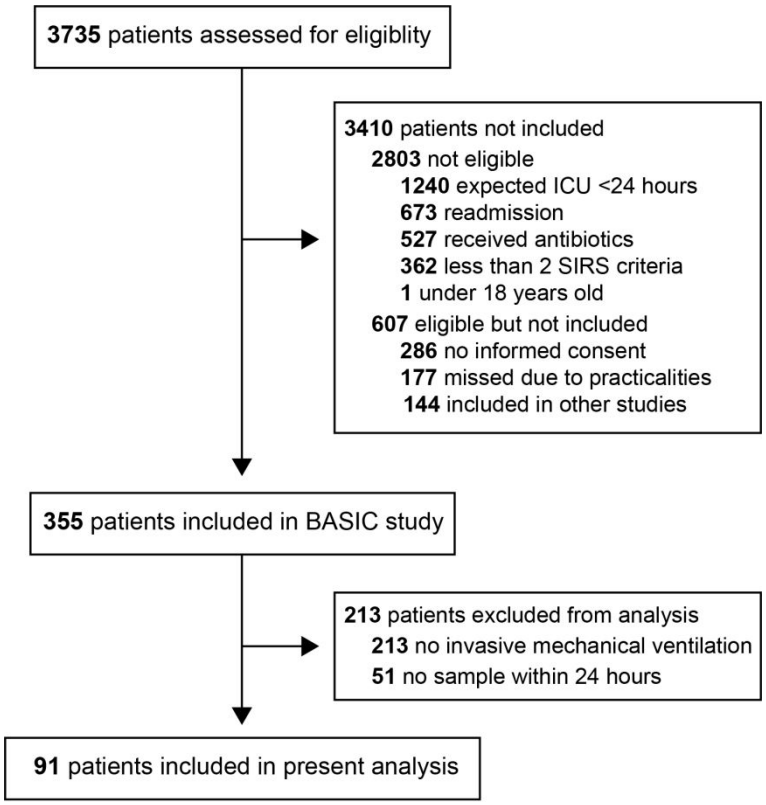
407 Sequences are available via the NCBI Sequence Read Archive (accession number
408 PRJNA553560). OTU tables, taxonomy classification tables, and metadata tables are available
409 at https://github.com/dicksonlunglab/MARS_lung_microbiome.

410

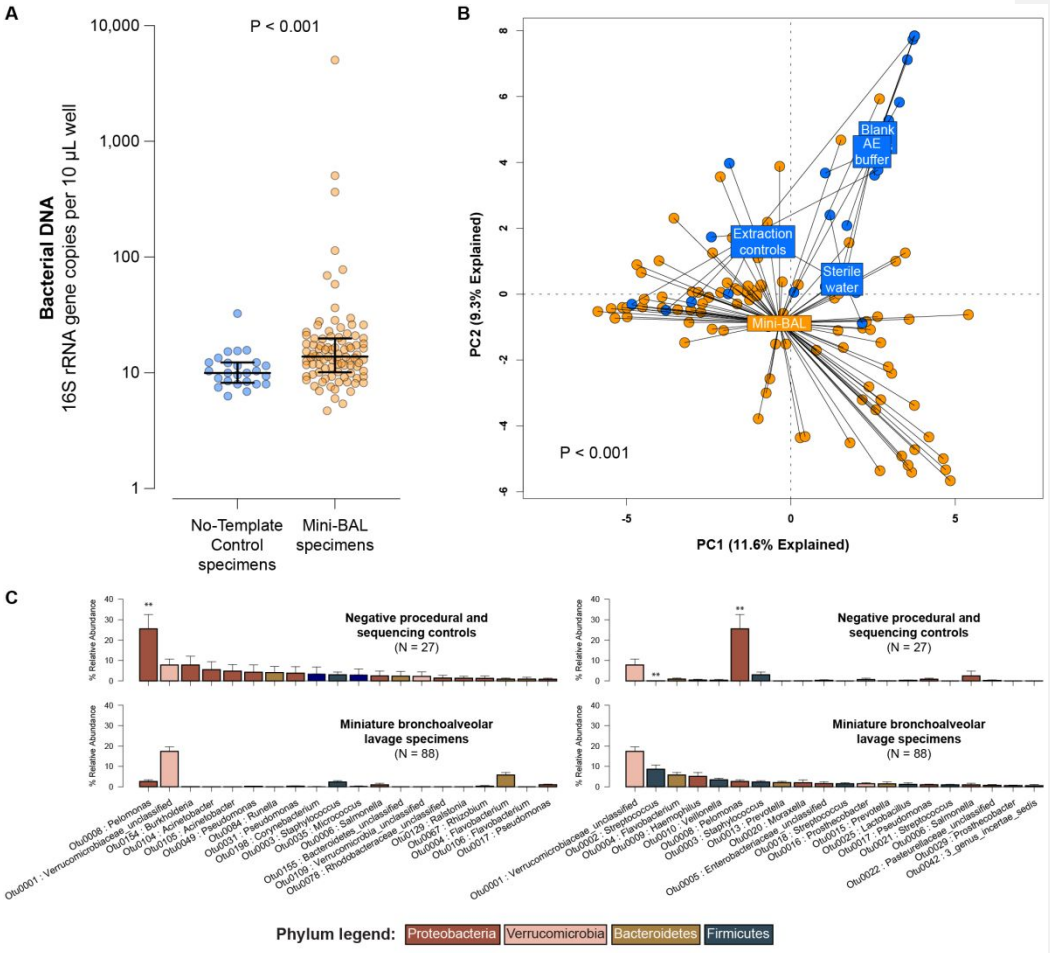
411 **Acknowledgements**

412 The content is solely the responsibility of the authors and does not necessarily represent the
413 official views of the National Institutes of Health. The authors thank Carolyn Calfee, Ariane
414 Panzer, and Susan Lynch for sharing the representative sequence of
415 OTU2119418:*Enterobacteriaceae*.

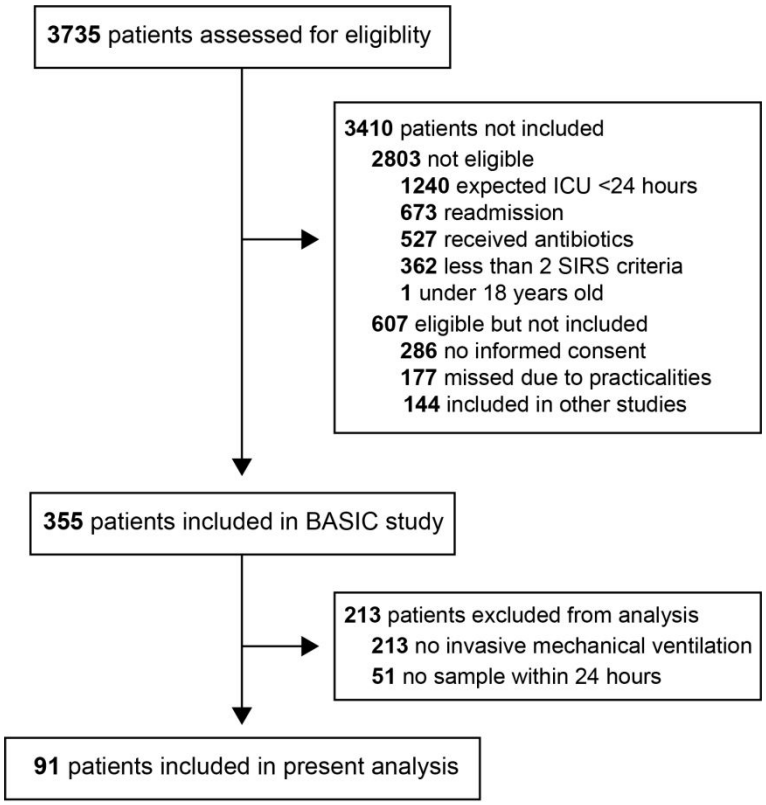
416



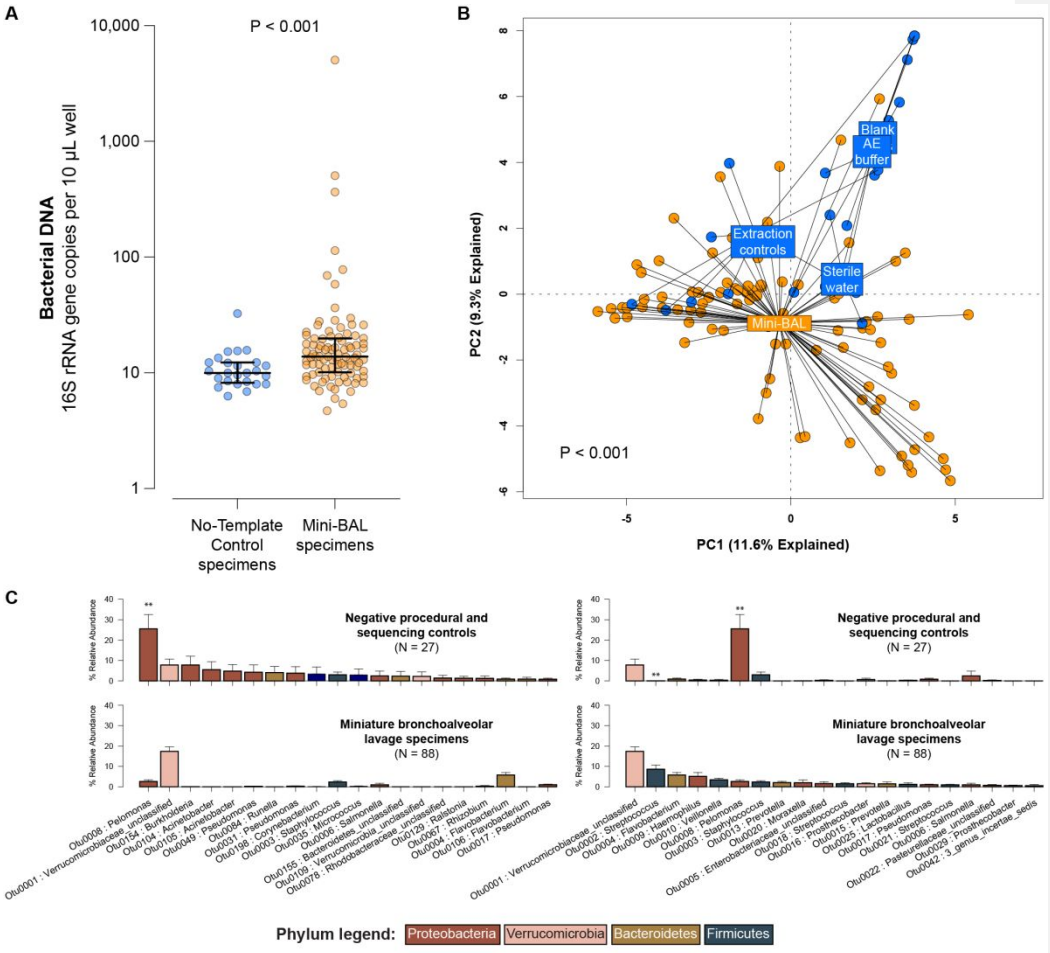
Supplemental Figure 1. CONSORT flow diagram of patient screening, inclusion, and analysis. The BASIC study (Biomarker Analysis in Septic Intensive Care patients) was a prospective, observational cohort study, of which the current study is a sub-analysis.



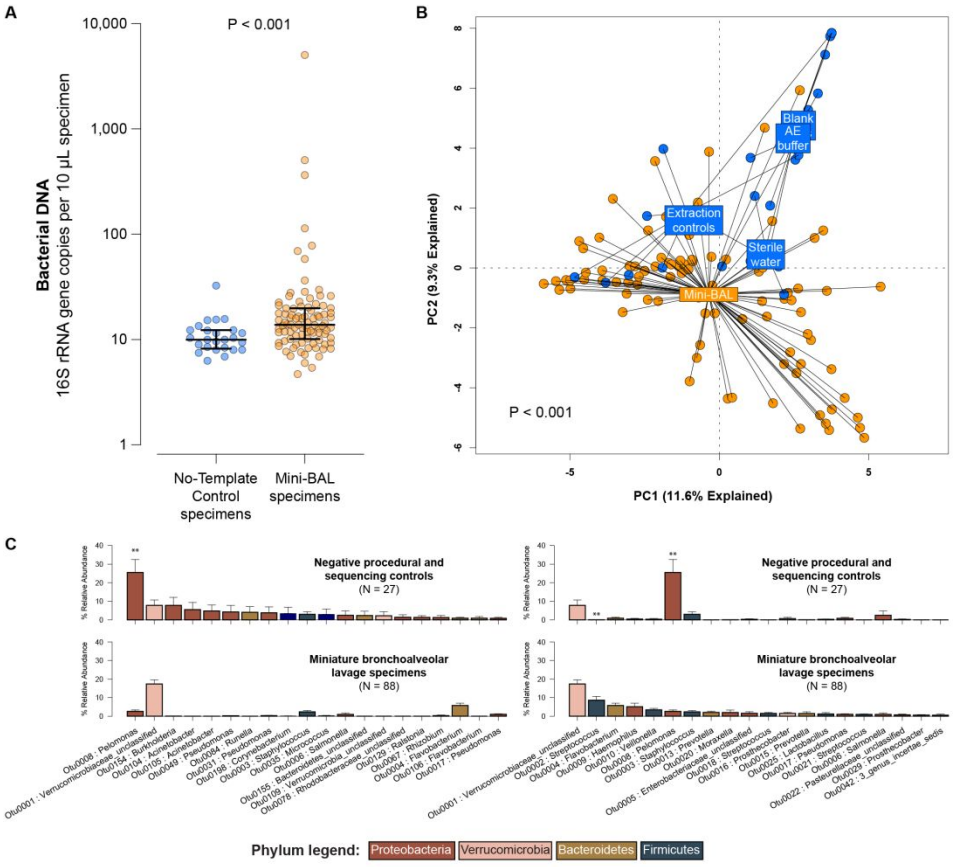
Supplemental Figure 2. Evidence of distinct bacterial signal in mini-bronchoalveolar lavage specimens compared to negative control specimens. Bacterial DNA in miniature-bronchoalveolar lavage (mini-BAL) specimens was quantified and identified using droplet digital PCR (A) and 16S rRNA gene amplicon sequencing (B, C), respectively. (A) The bacterial burden of mini-BAL specimens, while highly variable, was significantly greater than that of negative control specimens. (B) The community composition of bacteria identified in mini-BAL specimens was significantly distinct from that of bacteria detected in negative control specimens. (C) Rank abundance comparison of prominent taxa detected in negative controls (left) and mini-BAL (right) specimens. For each comparison, the twenty most abundant taxa in each group are displayed in decreasing order of relative abundance (mean ± SD). Asterisks indicated taxa idendified as significantly distinct across groups. Hypothesis testing performed using Mann-Whitney U test (A) and *mvabund* (B,C). A: median and interquartile range. B: mean and standard deviation.



Supplemental Figure 1. CONSORT flow diagram of patient screening, inclusion, and analysis. The BASIC study (Biomarker Analysis in Septic Intensive Care patients) was a prospective, observational cohort study, of which the current study is a sub-analysis.



Supplemental Figure 2. Evidence of distinct bacterial signal in mini-bronchoalveolar lavage specimens compared to negative control specimens. Bacterial DNA in miniature-bronchoalveolar lavage (mini-BAL) specimens was quantified and identified using droplet digital PCR (A) and 16S rRNA gene amplicon sequencing (B, C), respectively. (A) The bacterial burden of mini-BAL specimens, while highly variable, was significantly greater than that of negative control specimens. (B) The community composition of bacteria identified in mini-BAL specimens was significantly distinct from that of bacteria detected in negative control specimens. (C) Rank abundance comparison of prominent taxa detected in negative controls (left) and mini-BAL (right) specimens. For each comparison, the twenty most abundant taxa in each group are displayed in decreasing order of relative abundance (mean ± SD). Asterisks indicated taxa idendified as significantly distinct across groups. Hypothesis testing performed using Mann-Whitney U test (A) and *mvabund* (B,C). A: median and interquartile range. B: mean and standard deviation.



Supplemental Figure 1. Evidence of distinct bacterial signal in mini-bronchoalveolar lavage specimens compared to negative control specimens. Bacterial DNA in miniature-bronchoalveolar lavage (mini-BAL) specimens was quantified and identified using droplet digital PCR (A) and 16S rRNA gene amplicon sequencing (B, C), respectively. (A) The bacterial burden of mini-BAL specimens, while highly variable, was significantly greater than that of negative control specimens. (B) The community composition of bacteria identified in mini-BAL specimens was significantly distinct from that of bacteria detected in negative control specimens. (C) Rank abundance comparison of prominent taxa detected in negative controls (left) and mini-BAL (right) specimens. For each comparison, the twenty most abundant taxa in each group are displayed in decreasing order of relative abundance (mean ± SD). Asterisks indicated taxa idendified as significantly distinct across groups. Hypothesis testing performed using Mann-Whitney U test (A) and *mvabund* (B,C). A: median and interquartile range. B: mean and standard deviation.

Figure 2. (A) In our study, the presence of ARDS was associated with enrichment of an operational taxonomic unit (OTU) classified as *Enterobacteriaceae* (Otu0005, orange). This acterial DNA was 96% aligned with an ARDS-associated *Enterobacteriaceae* OTU from a prior study of the lung microbiome in critically ill patients (Panzer *AJRCCM* 2018) (Otu2119418, For each *Enterobacteriaceae* OTU, we identified the 20 most closely-aligned sequences identified via the SILVA database project and compared these OTUs with their neighbors, each r prominent taxa from our data set (classified as *Streptococcus*, *Staphylococcus*, and *Haemophilus*).

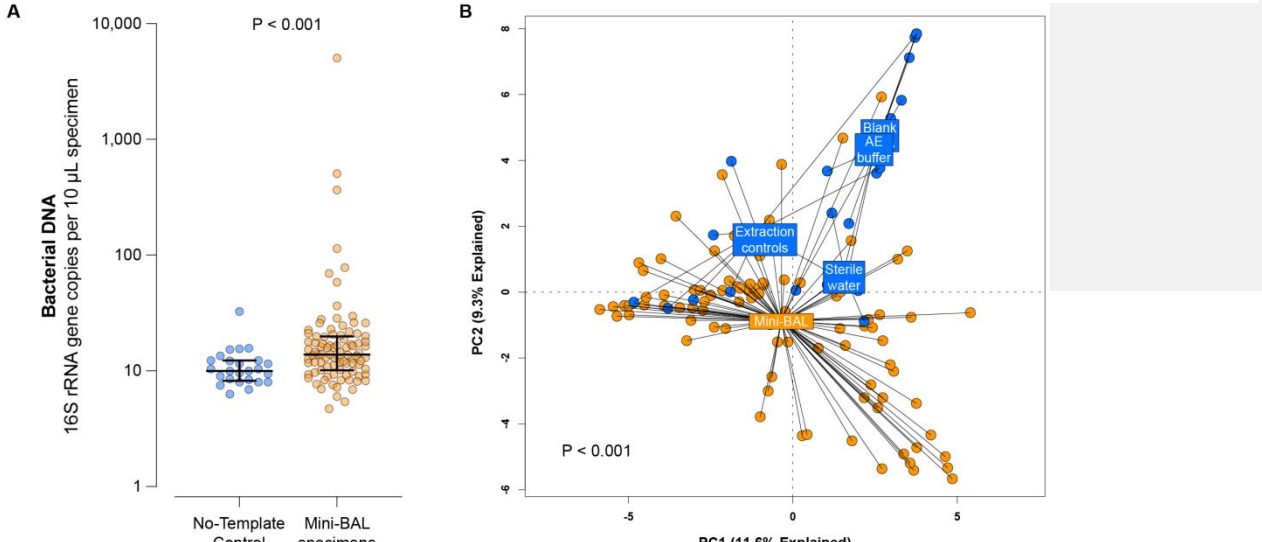


Table E1. Diversity Indices as Predictors of Ventilator-Free Days in Mechanically Ventilated Critically Ill Patients

Predictor	*Univariate*	
	Hazard ratio (CI)	*P* value
Shannon diversity index*	1.27 (0.87 – 1.86)	0.21
Community richness†	1.01 (0.98 – 1.00)	0.09
Community dominance‡	1.02 (1.00 – 1.04)	0.10

CI = Confidence interval (95%).
* per 1-unit increase.
† Unique OTUs per 1000 sequences, per OTU
‡ Relative abundance of most dominant OTU, per % relative abundance

Supplemental Figure 2. (A) In our study, the presence of ARDS was associated with enrichment of an operational taxonomic unit (OTU) classified as *Enterobacteriaceae* (Otu0005, orange). This sequence of bacterial DNA was 96% aligned with an ARDS-associated *Enterobacteriaceae* OTU from a prior study of the lung microbiome in critically ill patients (Panzer *AJRCCM* 2018) (Otu2119418, light blue). (B) For each *Enterobacteriaceae* OTU, we identified the 20 most closely-aligned sequences identified via the SILVA database project and compared these OTUs with their neighbors, each other, and other prominent taxa from our data set (classified as *Streptococcus*, *Staphylococcus*, and *Haemophilus*).

Commented [LB1]: Should change to supplemental figure 3

Table E1. Documented Admission Diagnoses of Study Cohort

Diagnosis	Study cohort (*n* = 91)
Cardiac arrest	19 (21)
Cerebral vascular accident (ischemic or hemorrhagic)	9 (10)
Pneumonia	7 (8)
Congestive heart failure/cardiomyopathy	6 (7)
Aortic aneurysm (rupture or dissection)	5 (5)
Sepsis	5 (5)
Meningitis	4 (4)
Pulmonary embolism	3 (3)
Bowel perforation	3 (3)
Gastrointestinal hemorrhage	2 (2)
Other	28 (31)

"Other" includes: arrhythmia, cholangitis, elective surgery, hemoptysis, hypoglycemia, obtundation, pyelonephritis, rhabdomyolysis, seizures, soft tissue infection, trauma.

Table E2. Diversity Indices as Predictors of Ventilator-Free Days in Mechanically Ventilated Critically Ill Patients

Predictor	*Univariate*	
	Hazard ratio (CI)	*P* value
Shannon diversity index*	1.27 (0.87 - 1.86)	0.21
Community richness†	1.01 (0.98 - 1.00)	0.09
Community dominance‡	1.02 (1.00 - 1.04)	0.10

CI = Confidence interval (95%).
* per 1-unit increase.
† Unique OTUs per 1000 sequences, per OTU
‡ Relative abundance of most dominant OTU, per % relative abundance

References

1. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The microbiome and the respiratory tract. *Annu Rev Physiol* 2016; 78: 481-504.
2. Dickson RP, Erb-Downward JR, Falkowski NR, Hunter EM, Ashley SL, Huffnagle GB. The lung microbiota of healthy mice are highly variable, cluster by environment, and reflect variation in baseline lung innate immunity. *Am J Respir Crit Care Med* 2018; 198: 497-508.
3. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, Curtis JL. Bacterial topography of the healthy human lower respiratory tract. *MBio* 2017; 8.
4. Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG, Li Y, Shen N, Ghedin E, Morris A, Diaz P, Huang L, Wikoff WR, Ubeda C, Artacho A, Rom WN, Sterman DH, Collman RG, Blaser MJ, Weiden MD. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat Microbiol* 2016; 1: 16031.
5. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. *Lancet* 2014; 384: 691-702.
6. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt MF, Cookson WO. Disordered microbial communities in asthmatic airways. *PLoS One* 2010; 5: e8578.
7. Dickson RP, Singer BH, Newstead MW, Falkowski NR, Erb-Downward JR, Standiford TJ, Huffnagle GB. Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nat Microbiol* 2016; 1: 16113.
8. Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, Murphy E, Johnston SL, Schwartz DA, Wells AU, Cookson WO, Maher TM, Moffatt MF. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 190: 906-913.

9. O'Dwyer DN, Ashley SL, Gurczynski SJ, Xia M, Wilke C, Falkowski NR, Norman KC, Arnold KB, Huffnagle GB, Salisbury ML, Han MK, Flaherty KR, White ES, Martinez FJ, Erb-Downward JR, Murray S, Moore BB, Dickson RP. Lung microbiota contribute to pulmonary inflammation and disease progression in pulmonary fibrosis. *Am J Respir Crit Care Med* 2019.

10. Sibila O, Laserna E, Shoemark A, Keir HR, Finch S, Rodrigo-Troyano A, Perea L, Lonergan M, Goeminne PC, Chalmers JD. Airway bacterial load and inhaled antibiotic response in bronchiectasis. *Am J Respir Crit Care Med* 2019; 200: 33-41.

11. Leitao Filho FS, Alotaibi NM, Ngan D, Tam S, Yang J, Hollander Z, Chen V, FitzGerald JM, Nislow C, Leung JM, Man SFP, Sin DD. Sputum Microbiome Is Associated with 1-Year Mortality after Chronic Obstructive Pulmonary Disease Hospitalizations. *Am J Respir Crit Care Med* 2019; 199: 1205-1213.

12. Rogers GB, Zain NM, Bruce KD, Burr LD, Chen AC, Rivett DW, McGuckin MA, Serisier DJ. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Ann Am Thorac Soc* 2014; 11: 496-503.

13. Bosch A, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Biesbroek G, Kool J, Pernet P, de Groot PCM, Eijkemans MJC, Keijser BJF, Sanders EAM, Bogaert D. Maturation of the infant respiratory microbiota, environmental drivers, and health consequences. A prospective cohort study. *Am J Respir Crit Care Med* 2017; 196: 1582-1590.

14. Kelly BJ, Imai I, Bittinger K, Laughlin A, Fuchs BD, Bushman FD, Collman RG. Composition and dynamics of the respiratory tract microbiome in intubated patients. *Microbiome* 2016; 4: 7.

15. Panzer AR, Lynch SV, Langelier C, Christie JD, McCauley K, Nelson M, Cheung CK, Benowitz NL, Cohen MJ, Calfee CS. Lung microbiota is related to smoking status and to development of ARDS in critically ill trauma patients. *Am J Respir Crit Care Med* 2017; 197: 621-631.

16. Zakharkina T, Martin-Loeches I, Matamoros S, Povoa P, Torres A, Kastelijn JB, Hofstra JJ, de Wever B, de Jong M, Schultz MJ, Sterk PJ, Artigas A, Bos LDJ. The dynamics of the pulmonary microbiome during mechanical ventilation in the intensive care unit and the association with occurrence of pneumonia. *Thorax* 2017; 72: 803-810.

17. Scicluna BP, Klein Klouwenberg PM, van Vught LA, Wiewel MA, Ong DS, Zwinderman AH, Franitza M, Toliat MR, Nurnberg P, Hoogendijk AJ, Horn J, Cremer OL, Schultz MJ, Bonten MJ, van der Poll T. A molecular biomarker to diagnose community-acquired pneumonia on intensive care unit admission. *Am J Respir Crit Care Med* 2015; 192: 826-835.

18. Klein Klouwenberg PM, Ong DS, Bos LD, de Beer FM, van Hooijdonk RT, Huson MA, Straat M, van Vught LA, Wieske L, Horn J, Schultz MJ, van der Poll T, Bonten MJ, Cremer OL. Interobserver agreement of Centers for Disease Control and Prevention criteria for classifying infections in critically ill patients. *Crit Care Med* 2013; 41: 2373-2378.

19. van Vught LA, Klein Klouwenberg PM, Spitoni C, Scicluna BP, Wiewel MA, Horn J, Schultz MJ, Nurnberg P, Bonten MJ, Cremer OL, van der Poll T. Incidence, Risk Factors, and Attributable Mortality of Secondary Infections in the Intensive Care Unit After Admission for Sepsis. *Jama* 2016; 315: 1469-1479.

20. Scicluna BP, van Vught LA, Zwinderman AH, Wiewel MA, Davenport EE, Burnham KL, Nurnberg P, Schultz MJ, Horn J, Cremer OL, Bonten MJ, Hinds CJ, Wong HR, Knight JC, van der Poll T. Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study. *Lancet Respir Med* 2017; 5: 816-826.

21. Bos LD, Schouten LR, van Vught LA, Wiewel MA, Ong DSY, Cremer O, Artigas A, Martin-Loeches I, Hoogendijk AJ, van der Poll T, Horn J, Juffermans N, Calfee CS, Schultz MJ. Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis. *Thorax* 2017; 72: 876-883.

22. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; 34: 1297-1310.

23. Jones AE, Trzeciak S, Kline JA. The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med* 2009; 37: 1649-1654.

24. Mason KL, Erb Downward JR, Mason KD, Falkowski NR, Eaton KA, Kao JY, Young VB, Huffnagle GB. Candida albicans and bacterial microbiota interactions in the cecum during recolonization following broad-spectrum antibiotic therapy. *Infect Immun* 2012; 80: 3371-3380.

25. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 2011; 108 Suppl 1: 4516-4522.

26. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013; 79: 5112-5120.

27. Oksanen JF, Blanchet G, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. vegan: Community Ecology Package. R package version 2.0-4.; 2012.

28. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

29. Wang Y, Naumann U, Wright ST, Warton DI. mvabund- an R package for model-based analysis of multivariate abundance data. *Methods Ecol Evol* 2012; 3: 471-474.

30. Schloss PD. MiSeq SOP - mothur. 2015. Available from: http://www.mothur.org/wiki/MiSeq_SOP.

31. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009; 75: 7537-7541.

32. Liaw A, Wiener M. Classification and Regression by randomForest. *R Sews* 2002; 2: 18-22.

33. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014; 12: 87.

34. Dickson RP, Erb-Downward JR, Prescott HC, Martinez FJ, Curtis JL, Lama VN, Huffnagle GB. Cell-associated bacteria in the human lung microbiome. *Microbiome* 2014; 2: 28.

35. Schuijt TJ, Lankelma JM, Scicluna BP, de Sousa e Melo F, Roelofs JJ, de Boer JD, Hoogendijk AJ, de Beer R, de Vos A, Belzer C, de Vos WM, van der Poll T, Wiersinga WJ. The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut* 2016; 65: 575-583.

36. Emonet S, Lazarevic V, Leemann Refondini C, Gaia N, Leo S, Girard M, Nocquet Boyer V, Wozniak H, Despres L, Renzi G, Mostaguir K, Dupuis Lozeron E, Schrenzel J, Pugin J. Identification of respiratory microbiota markers in ventilator-associated pneumonia. *Intensive Care Med* 2019.

37. Dickson RP, Erb-Downward JR, Huffnagle GB. Towards an ecology of the lung: new conceptual models of pulmonary microbiology and pneumonia pathogenesis. *Lancet Respir Med* 2014; 2: 238-246.

38. Schurink CAM, Nieuwenhoven CAV, Jacobs JA, Rozenberg-Arska M, Joore HCA, Buskens E, Hoepelman AIM, Bonten MJM. Clinical pulmonary infection score for ventilator-associated pneumonia: accuracy and inter-observer variability. *Intensive Care Med* 2004; 30: 217-224.