Supplementary information

Materials and methods

DNA extraction and 16S rRNA gene sequencing

We had 15 negative controls. They were the reagents that were used for the DNA extraction without containing any samples. DNA was extracted from 500 ul BALF and negative control using the Maxwell® RSC Whole Blood DNA Kit (Promega, USA). Each sample was centrifuged for 5 minutes at 14,000g to form a pellet. The supernatant was discarded and pellet was re-suspended in 720 ul of PBS with 80 ul proteinase K (Tiangen, China) added. Cells were lysed by bead beating using a FastPrep system (Qbiogene, Carlsbad, CA) for 3 cycles of 30 seconds at 6.0 m/sec and the mixture was subsequently incubated at 56 °C for 1 hour. Then, DNA was extracted according to the manufacturer's instructions. Extracted DNA was quantified using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States) and stored at -20 °C. The V3-V4 hypervariable region of the 16S rRNA gene of all samples was amplified by polymerase chain reaction (PCR) in 25-ul reactions. The amplicon primers were Forward Primer =5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG and Reverse Primer=5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC. The PCR conditions were 95°C for 3 min; followed by 25 cycles of 95°C for 30 s, 56°C for 30 s, and 72°C for 30 s; and then 72°C for 5 min. Index PCR was the 2nd stage PCR step that attached dual indices and Illumina sequencing adapters using the Nextera XT Index Kit, Index Primer 1(N7xx) and Index Primer 2(S5xx) (Illumina, San Diego, CA) in 50 ul reaction.

The PCR conditions were 95℃ for 3 min, 12 cycles at 95℃ for 30 s, 55℃ for 30 s, and 72℃ for 30 s, then 72℃ for 5 min. The amplicons were purified using Agencourt AMPure XP (Beckman Coulter) and quantified using KAPA Library Quantification Kit (KAPA Biosystems). The pooled library was sequenced on an Illumina Miseq platform (Illumina, San Diego, CA, United States) using pair-end sequencing (2×300bp).

Statistical analysis

The software VSEARCH (version 2.7.1) and USEARCH version (10.0) were used to process the sequencing data[1, 2]. Reads were denoised into Zero-radius Operational Taxonomic Units (ZOTUs) with UNOISE3[3]. After removal of ZOTUs identified as contaminants with decontam package or observed in the controls (Table S2) and whose relative abundance was less than 0.01%, a total of 490 ZOTUs were analyzed. Statistical analysis was performed in R version 3.6.2 via the Rstudio interface. We tested the association between the clinical factors and the lung microbiota composition of the patients using univariate PERMANOVA (vegan R-package) based on Bray-Curtis distance. Multivariable PERMANOVA was performed to assess the independent association of the factors. To identify the clinical factors-associated taxon, a random forest learning approach (randomForest R-package) was used and we determined variable importance using 1000 forests. The value of mean decrease in accuracy for classification was defined as the importance of the bacteria. Multivariable-adjusted Cox regression was performed to assess the association between lung microbiota and clinical improvements. P values were corrected using the false discovery procedure (q). P value less than 0.05 and q value less than 0.1 were considered significant. R packages ggplot2, corrplot, survminer were used for data visualization.

Contamination identification on sequencing results

Across these 67 BALF samples, the total number of merged reads was 6,804,070 (86,359±41,631). We used decontam package to identify the contaminants Zero-radius Operational Taxonomic Units (ZOTUs) in our sequencing data[4]. The methods including “frequency”, “prevalence”, and “both” were performed individually. A total of 154 ZOTUs were considered as contaminants. 29.87% of the ZOTUs belonged to phylum *Proteobacteria* and 11.04% belonged to phylum *Firmicutes*. Then, we observed the relative abundance ZOTU2 and ZOTU3 that belonged to phylum *Proteobacteria* family *Oxalobacteraceae* genus *Massilia* were as high as 18.43% (percentage of all samples sequences). Family *Oxalobacteraceae* was common in soil and gut microbiota while had a low relative abundance in the lung microbiota[5-7]. Given that the two ZOTUs accounted for 41.15% of all sequences detected in negative controls, we removed them from our analysis. After removal of ZOTUs identified as contaminants with decontam package or observed in the controls and whose relative abundance was less than 0.01%, a total of 4,923,448 reads (55,626±41,681.5, 490 ZOTUs) were analyzed and all the samples were normalized to 15,000 reads

References

1. Rognes T, Flouri T, Nichols B, Quince C, Mahe F: VSEARCH: a versatile open source tool for metagenomics. PEERJ 2016, 4:e2584.

2. Edgar RC: UPARSE: highly accurate OTU sequences from microbial amplicon reads. NAT METHODS 2013, 10(10):996-998.

3. Edgar RC: UNOISE2: Improved Error-Correction for Illumina 16S and ITS Amplicon Sequencing. bioRxiv 2016.

4. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ: Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. MICROBIOME 2018, 6(1):226.

5. Dickson RP, Schultz MJ, van der Poll T, Schouten LR, Falkowski NR, Luth JE, Sjoding MW, Brown CA, Chanderraj R, Huffnagle GB et al: Lung Microbiota Predict Clinical Outcomes in Critically Ill Patients. AM J RESP CRIT CARE 2020.

6. Shi YC, Cai ST, Tian YP, Zhao HJ, Zhang YB, Chen J, Ren RR, Luo X, Peng LH, Sun G et al: Effects of Proton Pump Inhibitors on the Gastrointestinal Microbiota in Gastroesophageal Reflux Disease. Genomics Proteomics Bioinformatics 2019, 17(1):52-63.

7. Vuko M, Cania B, Vogel C, Kublik S, Schloter M, Schulz S: Shifts in reclamation management strategies shape the role of exopolysaccharide and lipopolysaccharide-producing bacteria during soil formation. MICROB BIOTECHNOL 2020, 13(2):584-598.

Table S1. Criteria for severe community-acquired pneumonia

Minor criteria
Respiratory rate ≥ 30 breaths/min
PaO_2/FiO_2 ratio ≤ 250 mmHg
Multilobar infiltration
Confusion/disorientation
Uremia (Blood Urea Nitrogen) level, 7.14 mmol/L)
Leukopenia (White blood cell count, $<4\times10^9$/L)
Thrombocytopenia (platelet count, $<100\times10^9$/L)
Hypothermia (core temperature, <36°C
Hypotension requiring aggressive fluid resuscitation
Major criteria
Invasive mechanical ventilation (IMV)
Septic shock with the need for vasopressors

Note: Patients were diagnosed by presenting at least one major criterion or at least three minor criteria at the time of intensive care unit (ICU) admission.

Table S2. Contamination identification on sequencing results

Phylum	Number of ZOTUs	% of all the samples sequences
Proteobacteria[a]	48	20.12%
Firmicutes	17	0.104%
Actinobacteria	11	0.46%
Bacteroidetes	10	0.01%
Cyanobacteria	3	0.09%
Synergistetes	3	4.29×10^{-3}%
Unassigned	64	0.03%
Total	156	20.82%

a: Consisting of ZOTU2 and ZOTU3 belonging to family *Oxalobacteraceae*.

Table S3. Relationship between clinical parameters and the lung microbiota

Variable	R^2	P value[a]
Age, years (continuous)	0.014	0.538
Gender (Male/Female)	0.012	0.734
Smoking status (Current/Former/Non-smokers)	0.036	0.185
Drinking history	0.010	0.921
Immunocompromised status	0.014	0.493
Chronic respiratory disease	0.016	0.298
Sample season (Spring/Summer/Autumn/Winter)	0.041	0.64
Symptoms		
Fever	0.014	0.519
Cough	0.011	0.81
Sputum	0.021	0.118
Dyspnea	0.013	0.578
Confusion/disorientation	0.016	0.345
Laboratory findings on admission		
PH (continuous)	0.014	0.482
PaO_2/FiO_2 (continuous)	0.007	0.984
Creatinine >106 μmol/L	0.016	0.282
Procalcitonin (continuous)	0.029	0.026
Initial radiographic findings		
Diffuse bilateral pulmonary infiltration	0.011	0.808
Pleural effusion	0.011	0.785
Pathogen identified		
Bacteria	0.033	0.013
Virus	0.010	0.877
Atypical pathogens	0.011	0.765
Probable IFD	0.018	0.25
Complications at sampling		
ARDS	0.018	0.251
Sepsis	0.014	0.42
Septic shock	0.014	0.545
AKI	0.033	0.012
Acute cardiac insufficiency	0.016	0.298
Antibiotic used before sampling	0.042	0.566
Carbapenems		
Beta-lactams plus fluoroquinolones only		
Beta-lactams only		
Fluoroquinolones only		
Vancomycin[b]	0.013	0.545
Mechanical ventilation time	0.040	0.097
≥2 days		
<2 days		
Never		
Severity variables at sampling		
APACHE-II (continuous)	0.014	0.457
PSI risk class (I/II/III/IV/V)	0.050	0.822
CURB-65 risk score (continuous)	0.015	0.4
7-category ordinal scale at sampling (class 5/6)	0.016	0.362
Days from illness onset to admission	0.026	0.687
1-3 days		
4-7 days		
≥8 days		
Hours from admission to sampling	0.049	0.018
<12 hours		
12-24 hours		
24-48 hours		
Plasma Biomarker on admission, pg/mL (continuous)		
IL-4	0.010	0.742
IL-6	0.012	0.625

IL-8	0.010	0.711
MIP-1beta	0.029	0.03
VEGF-A	0.007	0.989
MMP-9	0.007	0.964

a: The analysis was performed by PERMANOVA based on Bray-Curtis distance.

b: Consisting of vancomycin, teicoplanin and linezolid.

Abbreviation: IFD, invasive fungal disease; ARDS, acute respiratory distress syndrome; AKI, acute kidney injury.

Table S4. Association of lung microbiota with severity of pneumonia

	ARDS			Shock			CURB-65			PSI class		
Variable	**Present (n=29)**	**Absent (n=38)**	**P[a]**	**Present (n=21)**	**Absent (n=46)**	**P**	**≥3 (n=26)**	**<3 (n=41)**	**P**	**IV-V (n=41)**	**I-III (n=26)**	**P**
Pathogen identified			0.02			0.19			0.47			0.43
Bacteria only	3(10.34)	10(26.32)		5(23.81)	8(17.39)		3(11.54)	10(24.39)		8(19.51)	5(19.23)	
Virus-bacteria co-infection	2(6.90)	4(10.53)		4(19.05)	2(4.35)		3(11.54)	3(7.32)		3(7.32)	3(11.54)	
Virus only	17(58.62)	8(21.05)		6(28.57)	19(41.30)		9(34.62)	16(39.02)		13(31.71)	12(46.15)	
Others	7(24.14)	16(42.11)		6(28.57)	17(36.96)		11(42.31)	12(29.27)		17(41.46)	6(23.08)	
Richness (continuous)	162.93±73.45	166.63±65.03	0.35	162.19±76.66	166.33±64.97	0.89	186.23±71.74	151.59±63.25	0.026	165±72.96	165.08±61.59	0.89
Richness (categorical)			0.70			0.69			0.11			0.97
≥200	8(27.59)	13(34.21)		7(33.33)	14(30.43)		12(46.15)	9(21.95)		13(31.71)	8(30.77)	
199-150	12(41.38)	12(31.58)		6(28.57)	18(39.13)		8(30.77)	16(39.02)		15(36.59)	9(34.62)	
<150	9(31.03)	13(34.21)		8(38.10)	14(30.43)		6(23.08)	16(39.02)		13(31.71)	9(34.62)	
Shannon (continuous)	3.2±1.74	3.18±1.79	0.95	3.29±1.87	3.14±1.72	0.74	3.64±1.67	2.9±1.77	0.09	3.31±1.76	3.00±1.76	0.45
Shannon (categorical)			0.74			0.72			0.52			0.84
>4.5	7(24.14)	11(28.95)		7(33.33)	11(23.91)		8(30.77)	10(24.39)		12(29.27)	6(23.08)	
4.5-2	15(51.72)	16(42.11)		9(42.86)	22(47.83)		13(50)	18(43.90)		18(43.90)	13(50)	
<2	7(24.14)	11(28.95)		5(23.81)	13(28.26)		5(19.23)	13(31.71)		11(26.83)	7(26.92)	
Prevotellaceae[b], (continuous)	2.32±4.61	1.63±3.79	0.34	2.46±5.07	1.69±3.68	0.60	1.53±2.45	2.18±4.94	0.89	1.48±3.48	2.63±5.00	0.46
Prevotellaceae, (categorical)			0.61			0.11			0.93			0.39
>2.4	6(20.69)	6(15.79)		5(23.81)	7(15.22)		5(19.23)	7(17.07)		6(14.63)	6(23.08)	
2.4-0.3	12(41.38)	13(34.21)		4(19.05)	21(45.65)		9(34.62)	16(39.02)		14(34.15)	11(42.31)	
<0.3	11(37.93)	19(50)		12(57.14)	18(39.13)		12(46.15)	18(43.90)		21(51.22)	9(34.62)	
Actinomycetaceae, (continuous)	1.25±4.99	0.64±2.67	0.84	0.28±0.80	1.19±4.58	0.19	0.43±0.87	1.21±4.84	0.35	0.88±4.20	0.95±3.25	0.89
Actinomycetaceae, (categorical)			0.99			0.69			0.61			0.91
>0.5	5(17.24)	7(18.42)		3(14.29)	9(19.57)		6(23.08)	6(14.63)		7(17.07)	5(19.23)	
0.5-0.005	12(41.38)	16(42.11)		8(38.10)	20(43.48)		11(42.31)	17(41.46)		18(43.90)	10(38.46)	
<0.005	12(41.38)	15(39.47)		10(47.62)	17(36.96)		9(34.62)	18(43.90)		16(39.02)	11(42.31)	
Moraxellaceae, (continuous)	5.3±18.12	2.65±10.77	0.39	4.59±11.51	3.44±15.59	0.76	2.10±6.03	4.87±17.73	0.88	2.89±10.51	5.23±19.08	0.89
Staphylococcaceae, (continuous)	2.33±11.06	5.77±13.65	0.48	1.21±3.21	5.68±14.93	0.54	2.83±11.67	5.20±13.25	0.36	2.21±6.66	7.54±18.21	0.68
Streptococcaceae, (continuous)	4.64±8.63	11.57±20.95	0.13	9.19±13.69	8.29±18.47	0.56	8.45±14.15	8.65±18.78	0.14	9.36±15.03	7.33±20.02	0.14
Pseudomonadaceae, (continuous)	41.18±33.78	28.87±30.98	0.10	35.5±38.23	33.61±30.1	0.82	29.31±31.45	37.30±33.25	0.39	30.77±31.4	39.61±34.21	0.33
Enterobacteriaceae, (continuous)	0.34±1.38	2.82±11.76	0.28	2.00±7.98	1.63±9.42	0.39	4.18±14.14	0.20±0.66	0.34	1.23±5.8	2.56±12.48	0.92

a: P values were from χ^2 test, Fisher's exact test, or Mann-Whitney U test.

b: The relative abundance (%) of the taxa at family level was analyzed continuously or by organizing their value into ranges. We turned the continuous variables into categorical variables by organizing their value into ranges. Firstly, according to the value of the variable, 67 samples were put in order. Secondly, we divided them into 10 groups evenly. Finally, we combined the neighboring groups that had the similar probability for the event happened.

Data were mean± (SD) for continuous variable and n (%) for categorical variable.

Figure legends

Figure S1. The distribution of patients in each changed category ordinal from Day 1 to 14 days.

D, day

Figure S2. Factors not associated with the lung microbiota composition

Principal Coordinate Analysis (PCoA) plot based on Bray-Curtis distance revealed that the lung microbial community composition was not significantly associated with Age (a), gender (b), baseline chronic respiratory diseases (c), immunocompromised status (d), season of sampling (e), PSI risk class (f), CURB-65 (g), plasma IL-6 level on admission (h), antibiotic use before bronchoscopy (i), detection of respiratory virus with 48 hours after admission (j), complicated with acute respiratory distress syndrome (ARDS k) and septic shock (l) at sampling.

Figure S3. The enrichment of Streptococcus spp in the lung microbiota was associated with high plasma MIP-1beta level

The positions of the ZOTUs on the Y-axis were determined by -log10 (P value from the comparison of the mean relative abundance in ZOTUs between the groups). The size of circles was proportional to the relative abundance of the ZOTUs in microbiota.

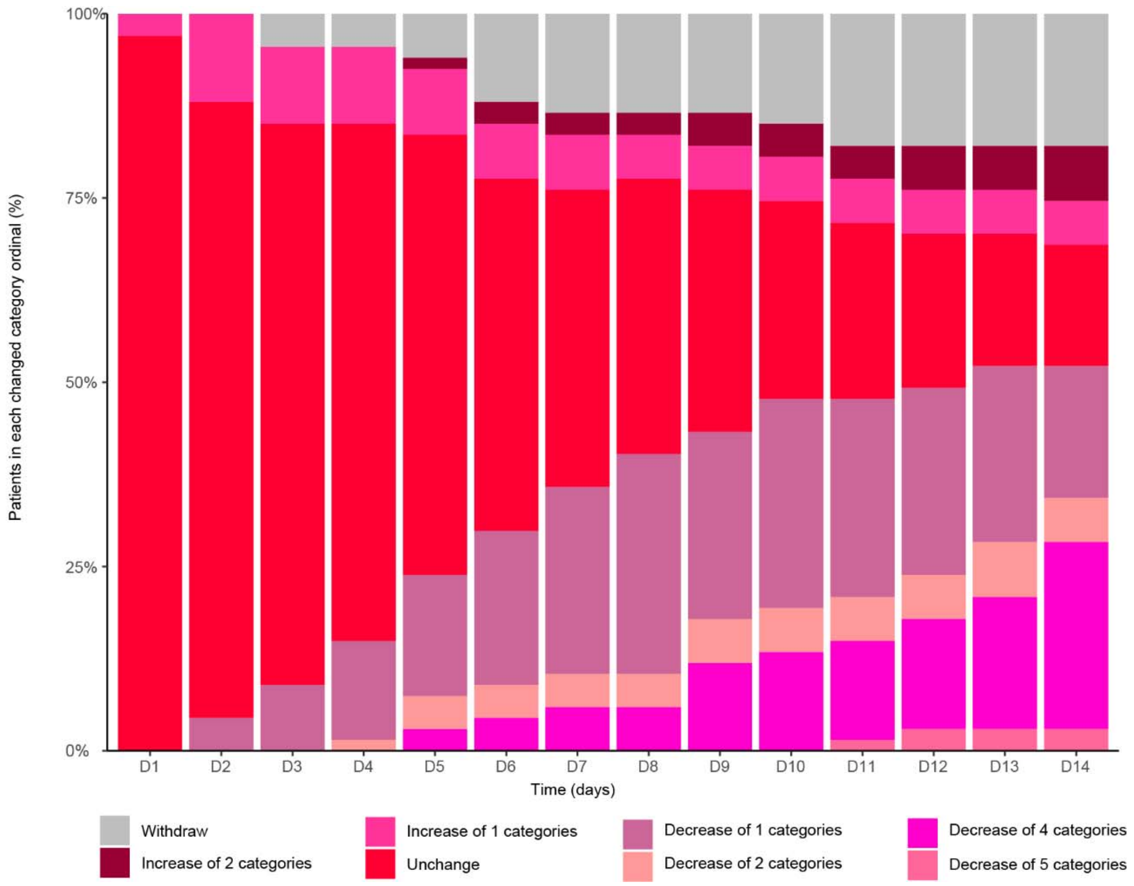
100%
75%
50%
25%
0%
Patients in each changed category ordinal (%)
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11
D12
D13
D14
Time (days)
Withdraw
Increase of 2 categories
Increase of 1 categories
Unchange
Decrease of 1 categories
Decrease of 2 categories
Decrease of 4 categories
Decrease of 5 categories

Figure S1.

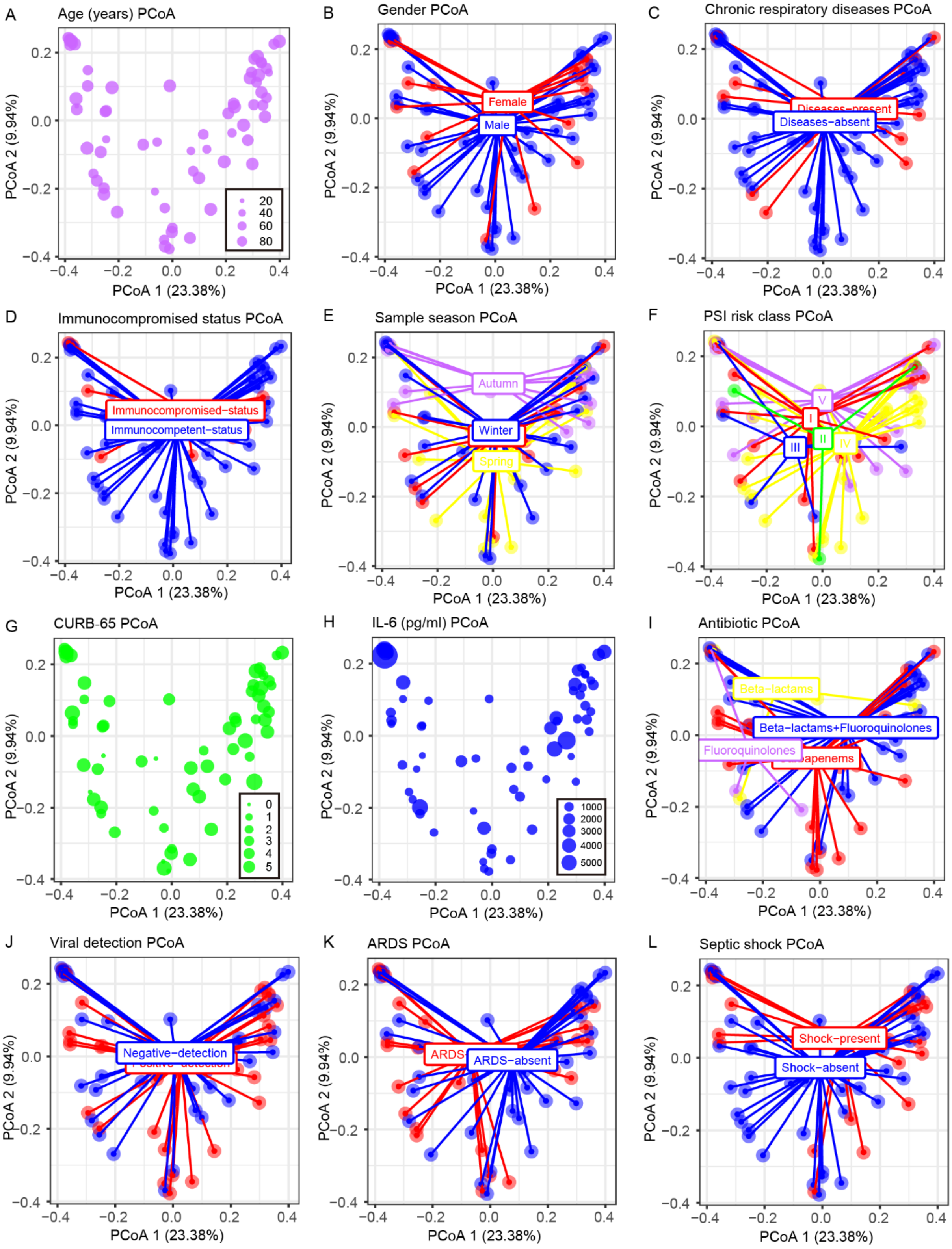
A
Age (years) PCoA
20
40
60
80
B
Gender PCoA
Female
Male
C
Chronic respiratory diseases PCoA
Diseases−present
Diseases−absent
D
Immunocompromised status PCoA
Immunocompromised−status
Immunocompetent−status
E
Sample season PCoA
Autumn
Winter
Spring
F
PSI risk class PCoA
V
I
II
III
IV
G
CURB-65 PCoA
0
1
2
3
4
5
H
IL-6 (pg/ml) PCoA
1000
2000
3000
4000
5000
I
Antibiotic PCoA
Beta−lactams
Beta−lactams+Fluoroquinolones
Fluoroquinolones
apenems
J
Viral detection PCoA
Negative−detection
K
ARDS PCoA
ARDS
ARDS−absent
L
Septic shock PCoA
Shock−present
Shock−absent
PCoA 2 (9.94%)
PCoA 1 (23.38%)
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4

Figure S2.

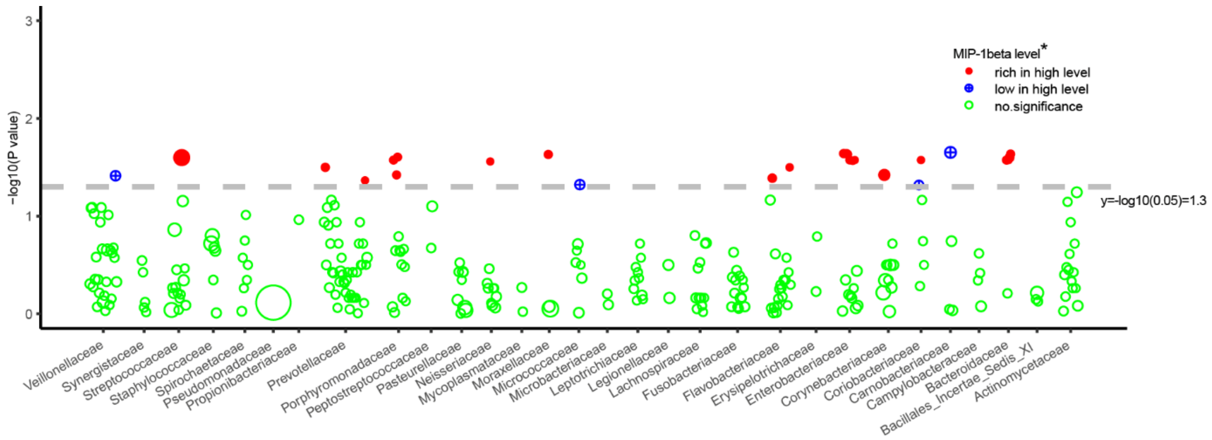
MIP-1beta level*
● rich in high level
⊕ low in high level
○ no.significance
−log10(P value)
3
2
1
0
y=-log10(0.05)=1.3
Veillonellaceae
Synergistaceae
Streptococcaceae
Staphylococcaceae
Spirochaetaceae
Pseudomonadaceae
Propionibacteriaceae
Prevotellaceae
Porphyromonadaceae
Peptostreptococcaceae
Pasteurellaceae
Neisseriaceae
Mycoplasmataceae
Moraxellaceae
Micrococcaceae
Microbacteriaceae
Leptotrichiaceae
Legionellaceae
Lachnospiraceae
Fusobacteriaceae
Flavobacteriaceae
Erysipelotrichaceae
Enterobacteriaceae
Corynebacteriaceae
Coriobacteriaceae
Carnobacteriaceae
Campylobacteraceae
Bacteroidaceae
Bacillales_Incertae_Sedis_XI
Actinomycetaceae

Figure S3