

1 **TITLE:** Pulmonary microbiome and transcriptome signatures reveal distinct pathobiologic states
2 associated with mortality in two cohorts of pediatric stem cell transplant patients.

3

4 **AUTHORS:** Matt S. Zinter^{1,2}, Christopher C. Dvorak², Madeline Y. Mayday^{1,3}, Gustavo Reyes¹, Miriam
5 R. Simon¹, Emma M. Pearce¹, Hanna Kim¹, Peter J. Shaw⁴, Courtney M. Rowan⁵, Jeffrey J. Auletta^{6,7}, Paul
6 L. Martin⁸, Kamar Godder⁹, Christine N. Duncan¹⁰, Nahal R. Lalefar¹¹, Erin M. Kreml¹², Janet R. Hume¹³,
7 Hisham Abdel-Azim^{14,15}, Caitlin Hurley¹⁶, Geoffrey D.E. Cuvelier¹⁷, Amy K. Keating^{18,10}, Muna Qayed¹⁹,
8 James S. Killinger²⁰, Julie C. Fitzgerald²¹, Rabi Hanna²², Kris M. Mahadeo^{23,8}, Troy C. Quigg^{24,25}, Prakash
9 Satwani²⁶, Paul Castillo²⁷, Shira J. Gertz^{28,29}, Theodore B. Moore³⁰, Benjamin Hanisch³¹, Aly Abdel-
10 Mageed²⁵, Rachel Phelan³², Dereck B. Davis³³, Michelle P. Hudspeth³⁴, Greg A. Yanik³⁵, Michael A.
11 Pulsipher³⁶, Imran Sulaiman^{37,38}, Leopoldo N. Segal³⁸, Birgitta A. Versluys^{39,40}, Caroline A. Lindemans^{38,39},
12 Jaap J. Boelens^{39,40,41}, Joseph L. DeRisi^{42,43}, on behalf of the *Pediatric Transplantation and Cell Therapy*
13 *Consortium**

14

15 **INSTITUTIONS:**

16 ¹ Division of Critical Care Medicine, Department of Pediatrics, University of California, San Francisco,
17 San Francisco, CA, USA

18 ² Division of Allergy, Immunology, and Bone Marrow Transplantation, Department of Pediatrics,
19 University of California, San Francisco, San Francisco, CA, USA

20 ³ Departments of Laboratory Medicine and Pathology, Yale School of Medicine, New Haven, CT, USA.

21 ⁴ The Children`s Hospital at Westmead, Sydney, Australia.

22 ⁵ Indiana University, Department of Pediatrics, Division of Critical Care Medicine, Indianapolis, IN, USA

23 ⁶ Hematology/Oncology/BMT and Infectious Diseases, Nationwide Children's Hospital, Columbus, OH,
24 USA.

25 ⁷ CIBMTR® (Center for International Blood and Marrow Transplant Research), National Marrow Donor
26 Program/Be The Match, Minneapolis, MN, USA.

27 ⁸ Division of Pediatric and Cellular Therapy, Duke University Medical Center, Durham, NC, USA.

28 ⁹ Cancer and Blood Disorders Center, Nicklaus Children's Hospital, Miami, FL, USA

29 ¹⁰ Harvard Medical School, Boston, Massachusetts; Division of Pediatric Oncology, Department of
30 Pediatrics, Dana-Farber Cancer Institute and Boston Children's Hospital, Boston, MA, USA.

31 ¹¹ Division of Pediatric Hematology/Oncology, UCSF Benioff Children's Hospital Oakland, University of
32 California San Francisco, Oakland, CA, USA.

33 ¹² Department of Child Health, Division of Critical Care Medicine, University of Arizona, Phoenix, AZ,
34 USA.

35 ¹³ University of Minnesota, Department of Pediatrics, Division of Critical Care Medicine, Minneapolis,
36 MN, USA

37 ¹⁴ Department of Pediatrics, Division of Hematology/Oncology and Transplant and Cell Therapy, Keck
38 School of Medicine, University of Southern California, Los Angeles, CA, USA

39 ¹⁵ Loma Linda University School of Medicine, Cancer Center, Children Hospital and Medical Center, Loma
40 Linda, CA, USA

41 ¹⁶ Division of Critical Care, Department of Pediatric Medicine, St Jude Children's Research Hospital,
42 Memphis, TN, USA

43 ¹⁷ CancerCare Manitoba, Manitoba Blood and Marrow Transplant Program, University of Manitoba,
44 Winnipeg, Manitoba, Canada

45 ¹⁸ Center for Cancer and Blood Disorders, Children's Hospital Colorado and University of Colorado,
46 Aurora, CO, USA.

47 ¹⁹ Aflac Cancer & Blood Disorders Center, Children's Healthcare of Atlanta and Emory University, Atlanta,
48 GA, USA.

49 ²⁰ Division of Pediatric Critical Care, Department of Pediatrics, Weill Cornell Medicine, New York, NY,
50 USA

51 ²¹ Department of Anesthesiology and Critical Care, Perelman School of Medicine, Children's Hospital of
52 Philadelphia, University of Pennsylvania, Philadelphia, PA, USA.

53 ²² Department of Pediatric Hematology, Oncology and Blood and Marrow Transplantation, Pediatric
54 Institute, Cleveland Clinic, Cleveland, OH, USA.

55 ²³ Department of Pediatrics, Division of Hematology/Oncology, MD Anderson Cancer Center, Houston,
56 TX, USA

57 ²⁴ Pediatric Blood and Marrow Transplantation Program, Texas Transplant Institute, Methodist Children's
58 Hospital, San Antonio, TX, USA.

59 ²⁵ Section of Pediatric BMT and Cellular Therapy, Helen DeVos Children's Hospital, Grand Rapids, MI,
60 USA.

61 ²⁶ Division of Pediatric Hematology, Oncology and Stem Cell Transplantation, Department of Pediatrics,
62 Columbia University, New York, NY, USA

63 ²⁷ University of Florida, Gainesville, UF Health Shands Children's Hospital, Gainesville, FL, USA

64 ²⁸ Department of Pediatrics, Division of Critical Care Medicine, Joseph M Sanzari Children's Hospital at
65 Hackensack University Medical Center, Hackensack, NJ, USA.

66 ²⁹ Department of Pediatrics, St. Barnabas Medical Center, Livingston, NJ, USA.

67 ³⁰ Department of Pediatric Hematology-Oncology, Mattel Children's Hospital, University of California, Los
68 Angeles, CA, USA.

69 ³¹ Children's National Hospital, Washington, District of Columbia, USA.

70 ³² Division of Pediatric Hematology/Oncology/BMT, Department of Pediatrics, Medical College of
71 Wisconsin, Milwaukee, WI, USA

72 ³³ Department of Pediatrics, Hematology/Oncology, University of Mississippi Medical Center, Jackson,
73 MS, USA.

74 ³⁴ Adult and Pediatric Blood & Marrow Transplantation, Pediatric Hematology/Oncology, Medical
75 University of South Carolina Children's Hospital/Hollings Cancer Center, Charleston, SC, USA.

76 ³⁵ Pediatric Blood and Bone Marrow Transplantation, Michigan Medicine, University of Michigan, Ann
77 Arbor, MI, USA.

78 ³⁶ Division of Hematology, Oncology, Transplantation, and Immunology, Primary Children's Hospital,
79 Huntsman Cancer Institute, Spense Fox Eccles School of Medicine at the University of Utah, Salt Lake
80 City, UT, USA.

81 ³⁷ Departments of Respiratory Medicine, Royal College of Surgeons in Ireland, Dublin, Ireland.

82 ³⁸ Department of Medicine, Division of Pulmonary and Critical Care Medicine, Laura and Isaac Perlmutter
83 Cancer Center, New York University Grossman School of Medicine, New York University (NYU) Langone
84 Health, New York, NY, USA.

85 ³⁹ Department of Stem Cell Transplantation, Princess Máxima Center for Pediatric Oncology, Utrecht,
86 Netherlands.

87 ⁴⁰ Division of Pediatrics, University Medical Center Utrecht, Utrecht, Netherlands

88 ⁴¹ Transplantation and Cellular Therapy, MSK Kids, Department of Pediatrics, Memorial Sloan Kettering
89 Cancer Center, New York, NY, USA.

90 ⁴² Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco,
91 CA, USA

92 ⁴³ Chan Zuckerberg Biohub, San Francisco, CA, USA

93

94 * PTCTC Investigators Listed in **eTable 1**.

95

96 **CORRESPONDENCE:** Matt Zinter, University of California, San Francisco; 1700 4th St, Byers Hall

97 404, San Francisco, CA 94158; matt.zinter@ucsf.edu

98

99 **AUTHOR CONTRIBUTIONS:** M.S.Z., C.C.D., G.A.Y., M.A.P., and J.L.D. contributed to the
100 conception and design of the work. All authors contributed to the acquisition of data. M.S.Z., C.C.D.,
101 M.Y.M., G.R., M.R.S., E.M.P., H.K., I.S., L.N.S., and J.L.D. contributed to the analysis of data. All
102 authors contributed to the drafting and revision of the manuscript.

103

104 **ACKNOWLEDGEMENTS:** M.S.Z. received research funding from NHLBI K23HL146936, NICHD
105 K12HD000850, the American Thoracic Society, the Pediatric Transplantation and Cell Therapy
106 Foundation, and the National Marrow Donor Program Amy Strelzer Manasevit Grant. M.Y.M received
107 research funding from NCI F31CA271571. H.A-A. received grant funding from the Gateway Foundation
108 and St. Baldrick’s Foundation. J.S.K. and J.J.B. received research funding from NCI P30CA008748.
109 M.A.P. received research funding from NCI P30CA040214. L.N.S. received research funding from NIGMS
110 R21GM147800, NCI R37CA244775, and NCI U2CCA271890. J.L.D. received research funding from the
111 Chan Zuckerberg Biohub. Additional funding for the study was provided by NHLBI UG1HL069254 and a
112 Johnny Crisstopher Children’s Charitable Foundation St. Baldrick’s Consortium Grant.

113

114 **DISCLOSURES:** M.S.Z. discloses consulting and advisory board work (Sobi). C.C.D. discloses
115 consulting and advisory board work (Jazz Pharmaceuticals; Alexion Inc.). J.J.A. discloses consulting and
116 advisory board work (AscellaHealth; Takeda). T.C.Q. discloses consulting and advisory board work
117 (Alexion AstraZeneca Rare Disease; Jazz Pharmaceuticals). H.A-A. discloses research support (Adaptive).
118 R.P. discloses consulting and advisory board work (BlueBird Bio) and research support (Amgen). M.A.P.
119 discloses consulting and advisory board work (Novartis; Pfizer; Cargo; BlueBird Bio; Vertex) and research
120 support (Miltenyi; Adaptive). L.N.S. discloses consulting and advisory board work (Sanofi). J.J.B. discloses
121 consulting and advisory board work (Sanofi; BlueRock; Sobi; SmartImmune; Immusoft; Advanced
122 Clinical; Merck). J.L.D. discloses salary support and research support (Chan Zuckerberg Biohub).

123 **ABSTRACT**

124 Lung injury is a major determinant of survival after pediatric hematopoietic cell transplantation (HCT). A
125 deeper understanding of the relationship between pulmonary microbes, immunity, and the lung epithelium
126 is needed to improve outcomes. In this multicenter study, we collected 278 bronchoalveolar lavage (BAL)
127 samples from 229 patients treated at 32 children's hospitals between 2014-2022. Using paired
128 metatranscriptomes and human gene expression data, we identified 4 patient clusters with varying BAL
129 composition. Among those requiring respiratory support prior to sampling, in-hospital mortality varied
130 from 22-60% depending on the cluster ($p=0.007$). The most common patient subtype, Cluster 1, showed a
131 moderate quantity and high diversity of commensal microbes with robust metabolic activity, low rates of
132 infection, gene expression indicating alveolar macrophage predominance, and low mortality. The second
133 most common cluster showed a very high burden of airway microbes, gene expression enriched for
134 neutrophil signaling, frequent bacterial infections, and moderate mortality. Cluster 3 showed significant
135 depletion of commensal microbes, a loss of biodiversity, gene expression indicative of fibroproliferative
136 pathways, increased viral and fungal pathogens, and high mortality. Finally, Cluster 4 showed profound
137 microbiome depletion with enrichment of Staphylococci and viruses, gene expression driven by lymphocyte
138 activation and cellular injury, and the highest mortality. BAL clusters were modeled with a random forest
139 classifier and reproduced in a geographically distinct validation cohort of 57 patients from The Netherlands,
140 recapitulating similar cluster-based mortality differences ($p=0.022$). Degree of antibiotic exposure was
141 strongly associated with depletion of BAL microbes and enrichment of fungi. Potential pathogens were
142 parsed from all detected microbes by analyzing each BAL microbe relative to the overall microbiome
143 composition, which yielded increased sensitivity for numerous previously occult pathogens. These findings
144 support personalized interpretation of the pulmonary microenvironment in pediatric HCT, which may
145 facilitate biology-targeted interventions to improve outcomes.

146

147 **BACKGROUND:**

148 Hematopoietic stem cell transplantation (HCT) involves high dose chemotherapy and/or radiation followed
149 by infusion of autologous or allogeneic hematopoietic progenitor cells with the intention of correcting
150 hematopoietic defects, rescuing chemotherapy-ablated marrow, or achieving a graft-versus-malignancy
151 effect.¹ HCT is often the only curative therapy for patients with life-limiting diseases such as malignancy,
152 bone marrow failure, and inborn errors of immunity, hemoglobin, and metabolism. However, direct
153 chemotherapy toxicity, opportunistic infection, and/or alloreactive inflammation can lead pulmonary injury
154 in up to 40% of patients,²⁻⁶ which can lead to hospital mortality rates approaching 50% when mechanical
155 ventilation is required.⁷⁻⁹

156 Given the severity of lung disease in this population, a deeper understanding of the pulmonary
157 microenvironment is needed to develop next-generation diagnostic tests and treatments that will improve
158 survival rates. The lung microenvironment is a complex interaction between pulmonary microbes,
159 immunity, and the lung epithelium and stroma, and significant questions regarding the role of pulmonary
160 microbes in relation to each other remain largely unanswered as they pertain to human health. We and
161 others have shown that the lungs are not sterile, and in fact contain a variety of microbes of varying
162 pathogenic potential that continually populate the lung due to inhalation, aspiration, and in some cases of
163 disease, hematogenous spread.¹⁰⁻¹² Lung sampling through bronchoscopic bronchoalveolar lavage (BAL)
164 is used clinically to detect common pathogens; however, many pathogens evade detection due to preceding
165 antimicrobial treatment, lack of serologic immunity in the post-HCT setting, or limited preselected targets
166 on multiplex assays, all of which may lead to delayed or missed diagnoses and prolonged broad-spectrum
167 antimicrobial exposure.^{13,14} In addition, organisms of indeterminate significance or context-dependent
168 virulence are frequently identified, leading to questions about the structure, composition, and significance
169 of broader microbial communities in this population.^{11,15}

170 We previously reported that in a cohort of children preparing to undergo allogeneic HCT, both pulmonary
171 microbial depletion and pathogen enrichment were associated with contemporaneously poor lung function,

172 concomitant inflammation, and the eventual development of fatal post-HCT lung disease.^{16,17} To expand
173 these findings to the post-HCT setting, we prospectively enrolled pediatric HCT patients undergoing
174 clinically-indicated BAL as part of evaluation for pulmonary complications. BAL underwent RNA
175 sequencing to characterize the pulmonary microbiome landscape, surveil for occult pulmonary infections,
176 and capture lung gene expression profiles. Overall, we found that depletion of commensal microbiome
177 constituents was associated with pathogen enrichment, acute inflammation, fibroproliferation, and poor
178 survival. We were able to distinguish common respiratory pathogens from commensals using a community-
179 structure analysis approach. Our results suggest a pathobiologic signature of dysbiotic lung injury that could
180 be adapted into next-generation diagnostics and eventually leveraged in new therapeutic pipelines to
181 improve outcomes.

182 **RESULTS:**

183 **Patients:** From 2014-2022, pediatric HCT recipients across 32 children's hospitals in the United States,
184 Canada, and Australia (**Figure 1A**) who developed pulmonary complications and were preparing to
185 undergo clinically-indicated bronchoscopic BAL were prospectively approached along with their
186 parents/guardians for research consent to cryopreserve unused BAL (**Figure 1B**). The final cohort included
187 n=278 BALs from n=229 patients (**Table 1**). Pulmonary symptoms developed or worsened a median 93
188 days after HCT (IQR 23-278) and were frequently associated with hypoxia and abnormal chest imaging,
189 often in the setting of other comorbidities such as GVHD and sepsis (**Table 2**). BAL was performed a
190 median 112 days after HCT (IQR 36-329), at which point lymphopenia was prevalent (median ALC 420
191 cells/uL, IQR 156-1,035, **eFigure 1**). Following each patient's most recent BAL procedure, 121/229
192 patients required intensive care (53%), 71 required ≥ 7 days of mechanical ventilation (31%), and 45 patients
193 died in the hospital (20%).

194 **Cluster Derivation:** BAL underwent mechanical homogenization, bulk RNA extraction, and sequencing,
195 followed by parallel alignment to microbial and human reference genomes using the open-source CZID
196 platform (czid.org) (**Figure 1C, Methods**).¹⁸ Microbial alignments were transformed from reads counts to

197 quantitative masses using a reference spike-in¹⁹, followed by stringent contamination subtraction²⁰, and
198 were summarized according to taxa, KEGG functional orthologs, richness, and diversity. Human
199 alignments were characterized according to normalized gene expression, pathway analysis, cell type
200 deconvolution, and T- and B-cell receptor alignments (**Methods**). We first used unsupervised analysis to
201 identify underlying BAL subtypes with shared microbial-human metatranscriptomic composition. We used
202 a two-step approach consisting of (1) multi-factor dimensionality reduction (*mofa*), followed by (2) uniform
203 manifold approximation and projection with hierarchical clustering (*umap*) to assess BAL compositional
204 similarity (**Methods**). Optimal fit statistics (**eFigures 2-4**) suggested that 4 clusters best fit the data structure
205 (**Figure 1D**).

206 **Clinical Traits, Illness Severity and Outcomes:** Clinical traits and outcomes were analyzed only after the
207 clusters were assigned. Demographics, medical disease, transplant regimens, and graft characteristics were
208 similar among clusters, with the exception of more females in Clusters 3 and 4 (**eTable 2**). However,
209 patients in Clusters 3 and 4 were generally sicker, as evidenced by greater need for respiratory support prior
210 to BAL ($p=0.004$), higher rates of renal injury and GVHD ($p=0.001$ and $p=0.019$), and greater use of
211 intensive care ($p=0.001$) or prolonged mechanical ventilation (≥ 7 days) after BAL ($p=0.001$, **eTable 3**).
212 Using each patient's most recent BAL, patients in Clusters 3 and 4 also had significantly higher in-hospital
213 mortality than patients in Clusters 1 or 2 (33 and 35% vs 14 and 14%, log-rank $p=0.005$, **Figure 1E**).
214 Among patients requiring respiratory support prior to BAL (44%), cluster-based mortality differences were
215 pronounced and ranged from 22-30% in Clusters 1 and 2 to 50-60% in Clusters 3 and 4 (log-rank $p=0.007$).
216 Findings were similar when analyzing only patients enrolled within 100 days post-HCT (**eTable 4**) and in
217 a multivariable Cox regression model accounting for age, biologic sex, ANC, ALC, and presence of GVHD
218 ($p=0.023$, **eTable 5**). Of note, only 2 patients died within 48 hours of BAL (both in the setting of
219 progressive septic shock).

220 **Microbial Taxonomy:** To determine how microbiome composition drove differences between the clusters,
221 we compared taxonomic mass, richness, and diversity. Cluster 1 was defined by moderate microbiome mass

222 and richness, high microbial diversity, and a low burden of viruses. In contrast, Cluster 2 showed high mass
223 of all bacterial phyla, as well as high levels of taxonomic richness and moderate microbial diversity (**Figure**
224 **2A, 2B, Data File 1**). Cluster 3 demonstrated a reduced quantity and diversity of typically oropharyngeal
225 microbes with greater quantity of RNA viruses and the Ascomycota phylum of fungi, which contains
226 medically-relevant pathogens such as *Aspergillus*, *Candida*, and *Pneumocystis*. In contrast, Cluster 4
227 showed significant depletion of typical microbiome constituents with minimal diversity and richness and
228 concomitant enrichment of *Staphylococcus* and the Pisuviricota phylum of RNA viruses, which contains
229 numerous respiratory RNA viruses such as Rhinovirus. BALs representative of each Cluster are shown in
230 **eFigure 5**. We next used an orthogonal supervised analysis to compare microbiome features among
231 survivors and non-survivors. Consistent with the description of Clusters 3 and 4, non-survivors showed
232 broad bacterial depletion of commensal taxa, higher quantities of fungal and viral RNA (**Figure 2C, Data**
233 **File 2**), and decreased BAL richness ($p=0.025$) and diversity (Shannon diversity $p=0.006$; **Figure 2D**). In
234 contrast, survivors showed replete and bacterially diverse pulmonary microbiomes, consistent with
235 description of Cluster 1.

236 **Microbial Function:** Transcriptomic markers of metabolic activity of microbial communities may
237 complement taxonomic composition.²¹ Therefore, we next characterized the 4 clusters according to KEGG
238 functional annotations. Cluster 1 showed moderate transcription of myriad microbial metabolic functions
239 across the domains of carbohydrate, lipid/fatty acid, and amino acid metabolism (**Figure 2E, eFigure 6,**
240 **Data File 3**). In contrast, the bacterially rich Cluster 2 showed greater transcription of these domains as
241 well as of glycan biosynthesis pathways, including peptidoglycan, lipopolysaccharide, and other glycans
242 that form bacterial cell walls (**eFigure 7**). Cluster 3 showed significantly lower microbial function across
243 the spectrum of KEGG pathways, and consistent with a depleted microbiome, Cluster 4 showed minimal
244 microbial metabolic activity. Select metabolic pathways are shown in **Figure 2F**. These results indicate that
245 functionally, the two clusters highly associated with poor outcome showed relative loss of common critical
246 microbial functions.

247 **Pathogen Identification:** Although some microbiome features were shared across clusters, such as the
248 degree of quantity of oropharyngeal taxa, many patients in this cohort had a wide range of distinct
249 infections, thus lending unique elements to each microbiome. Therefore, we characterized the landscape of
250 detected microbes with pathogenic potential relative to the clinical assay metadata from each patient
251 (summarized in **eTable 6**, pathogen list in **Data File 4**, patient-level data in **Data File 5**).

252 **Viruses:** Clinically, most community-acquired respiratory viruses (CRVs) are detected with multiplex PCR
253 and reported as present/absent. Clinical testing found CRVs in 18% of samples (n=49), whereas sequencing
254 identified CRVs in 28% of samples (n=77), highest in Clusters 2, 3, and 4 (**Figure 3A**). In addition to
255 common CRVs, several novel (< 90% nucleotide identity) or variant strains of common CRVs such as
256 Influenza C and Rhinovirus C were detected (GenBank OQ116581, OQ116582, OQ116583).²²⁻²⁴ Clinical
257 testing found herpesviruses (HVs) including CMV and HHV-6 in 13% of samples (n=35), whereas
258 sequencing found HVs in 16% of samples (n=49), with greatest detection in Clusters 3 and 4. Sequencing
259 also detected many viruses known to have respiratory transmission but not typically included on respiratory
260 viral panels, including BK, WU, and KI Polyomaviruses, Bocavirus, Parvovirus B19, lymphocytic
261 choriomeningitis virus (LCMV), and non-vaccine strain Rubella across 26 BALs from 23 patients. These
262 viruses were most common in Clusters 3 and 4 and associated with 39% in-hospital mortality (n=9/23). The
263 ubiquitous bystander torquetenovirus (TTV) and its variants were detected in 20% of samples (n=55), again
264 higher in Clusters 2, 3, and 4 relative to Cluster 1 (**eTable 7**, p<0.001).

265 **Bacteria:** Clinically, most pathogenic respiratory bacteria are detected with combination of selective culture
266 media (blood, chocolate, and McConkey agar) optimized to grow certain pathogens above non-pathogenic
267 background, although PCR, serology, and antigen tests may be used for certain organisms. In this study,
268 clinical testing identified pathogenic bacteria in 51 samples, which were heavily overrepresented in the
269 microbially-rich Cluster 2 (32 of 51 bacterial infections). In contrast, metagenomic sequencing is inherently
270 unbiased regardless of organism pathogenicity and thus can detect microbes broadly. Since contamination
271 is ubiquitous in low-biomass samples,^{25,26} we used a strict approach to adjust for background taxa using

272 internal spike-ins and a series of external controls (**Methods**).^{20,27} Still, many potentially pathogenic
273 microbes were detected broadly; for example, *S.pneumoniae*, *M.catarrhalis*, *H.influenzae*, *S.aureus*,
274 *P.aeruginosa* were detected in 34%, 21%, 21%, 16%, and 14% of samples (94, 58, 57, 44, and 39 samples),
275 respectively. Since some microbes could be present as commensals or pathogens, depending on context and
276 microbial burden, we then ranked bacteria according to RNA mass, dominance of the bacterial microbiome,
277 and intra-cohort z-score in order to parse microbes most likely to be present in states of dysbiosis and thus
278 potential infection (**Figure 3B**). Using a conservative threshold of RNA mass ≥ 10 pg, bacterial dominance
279 $\geq 20\%$, and Z-score $\geq +2$, we found potentially pathogenic bacteria in 76 samples, again with nearly half of
280 these in Cluster 2. In addition to new cases of common pathogens (e.g.: *P.aeruginosa*), numerous previously
281 occult pathogens were identified above these thresholds, including *B.cereus*, *C.freundii*, *C.pneumoniae*,
282 *K.aerogenes*, *S.enterica*, and *U.parvum*.

283 **Eukaryotes:** As with bacteria, many potentially pathogenic fungi were detected broadly in this cohort; for
284 example, Candida, Aspergillus, Fusarium, and Rhizopus were detected in 18%, 16%, 9%, and 5% of
285 samples (50, 44, 25, and 13), respectively. By clinical assays, potentially pathogenic fungi were detected in
286 9% of samples (n=25). Using sequencing with a threshold of mass ≥ 10 pg and Z-score $\geq +2$, potentially
287 pathogenic fungi were detected in 30% of samples (83), with high detection across clusters 2, 3, and 4
288 (**Figure 3C**). Several relevant fungi were detected exclusively by metagenomic sequencing, including
289 Cryptococcus and Pneumocystis. No BAL parasites were detected through clinical assays, whereas
290 metagenomic sequencing detected Toxoplasma in 4 patients and Acanthamoeba in 3 patients, with
291 predominance in Clusters 3 and 4 (**Data File 5**) and $>50\%$ mortality rate (n=4/7).

292 Overall, clinical testing identified 173 pathogens in 116/278 samples (41.7%), while metagenomic
293 sequencing using the above conservative thresholds identified 360 pathogens in n=196/278 samples
294 (70.5%, McNemar's $p < 0.001$, **eTable 8**). Combined, clinical testing and metagenomic sequencing together
295 identified 429 pathogens in n=209/278 samples (75.2%, **eTable 6**). Whereas clinical testing identified
296 pathogens in 22/45 non-survivors (49%), sequencing identified credible pathogens in 36/45 non-survivors

297 (80%, $p=0.002$). In-hospital mortality was highest for those with a pathogen detected by both clinical testing
298 and metagenomics and lower if a pathogen was detected by metagenomics alone or was not detected at all
299 (27% vs 19% vs 13%, **eTable 9, eFigure 8**).

300 **Impact of Antimicrobial Exposure:** Although the effects of antimicrobial exposure have been
301 demonstrated on the intestinal, nasal, and oropharyngeal microbiomes, the effects of antibiotics on the
302 bronchoalveolar microbiome are less clear, with some reporting a major effect²⁸⁻³⁵ and some reporting
303 minimal effect.^{36,37} To investigate this, we quantified patient-level antibacterial exposure in the week
304 preceding BAL by weighting the cumulative antibiotic exposure days (**Figure 4A**) with an agent-specific
305 broadness score³⁸ to yield an antibiotic exposure score (AES, **Figure 4B, Methods**). AES varied across
306 clusters ($p=0.005$) and was lowest for the microbially-rich Cluster 2 and highest for the microbially depleted
307 Clusters 3 and 4. Greater AES was associated with reduced BAL microbial richness (Spearman rho -0.14 ,
308 $p=0.018$); depletion of all the major bacterial phyla including numerous oropharyngeal-resident taxa; and
309 enrichment of the fungal phylum Ascomycota ($FDR<0.05$, **Figure 4C, Data File 6**). In addition, AES was
310 significantly greater among non-survivors (median 352, IQR 210-507 vs. 175, IQR 75-336, Wilcoxon rank-
311 sum $p<0.001$), with sequentially higher mortality with increasing AES quartile (**eFigure 9**). Using causal
312 mediation analysis based on linear structural equation modeling (**Methods**), the association between greater
313 AES and mortality was statistically mediated by an antibiotic-induced reduction in key commensal
314 pulmonary bacteria including *Actinomyces*, *Fusobacterium*, *Gemella*, *Haemophilus*, *Neisseria*, *Rothia*,
315 *Schaalia*, and *Streptococcus* ($p<0.001$), suggesting that the link between antibiotic exposure and mortality
316 can at least partially be explained by effects of antibiotics on the pulmonary microbiome (**eFigure 10, Data**
317 **File 7**). Many groups have found that anti-anaerobe exposure is associated with a depleted intestinal
318 microbiome and progression of upper respiratory viral infections to the lower respiratory tract.³⁹⁻⁴¹ Similar
319 to above, anti-anaerobic exposure was higher in non-survivors ($p=0.011$) and was associated with BAL
320 depletion of numerous anaerobes including *Prevotella*, *Gemella*, and *Fusobacterium* (**Data File 8**). Anti-
321 fungal exposure appeared higher in the microbially-depleted Cluster 4, driven largely by higher exposure

322 to echinocandins ($p=0.019$), and anti-viral exposure appeared higher in Clusters 3 and 4, driven largely by
323 higher exposure to cidofovir ($p=0.045$).

324 **Impact of Clinical Immune Status:** The pulmonary microbiome exists in a state of reciprocal interaction
325 with the lung epithelium, stroma, and immune system. To contextualize microbiome states according to
326 systemic immunity, we analyzed each patient's most recent blood absolute neutrophil count (ANC) and
327 absolute lymphocyte count (ALC) measured prior to BAL. ANC was highest in the bacterially-rich Cluster
328 2 ($p=0.029$, **eTable 3**) but was not associated with mortality overall ($p=0.810$). ALC did not vary across
329 clusters ($p=0.997$) but was significantly lower in non-survivors (median 273, IQR 125-650 vs. 422, IQR
330 179-1120, $p=0.028$).

331 **Pulmonary Gene Expression:** We then compared BAL human gene expression across the 4 clusters. A 4-
332 way ANOVA-like analysis yielded 18,158 genes differentially expressed across the 4 clusters (**Figure 5A**,
333 **Data File 9**). Select genes most differentially expressed in each cluster are displayed in **Figure 5B**. To
334 assess the biological pathways represented by these genes, we compared GSEA enrichment scores for
335 Reactome gene sets (**Data File 10**); select pathways most differentially expressed in each cluster are
336 displayed in **Figure 5C**. Overall, Cluster 1 showed high expression of pathways related to antigen-
337 presenting cell activation; Cluster 2 showed high expression of genes and pathways related to neutrophil
338 and innate immune activation, bacterial processing, and airway inflammation; Cluster 3 showed high
339 expression of pathways related to collagen deposition and fibroproliferation; and Cluster 4 showed high
340 expression of anti-viral and cellular injury genes. To replicate these findings using a different methodology
341 unrelated to the above clusters, we performed a supervised analysis comparing gene expression among
342 survivors and non-survivors and identified 1,253 differentially expressed genes (**Data File 11**). Consistent
343 with the description of Clusters 3 and 4, BALs from non-survivors showed broad down-regulation of innate
344 immune and antigen-presenting signals and a significant upregulation in collagen deposition, matrix
345 metalloproteinases, alveolar epithelial hyperplasia, and fibroproliferative genes (e.g.: COL1A1, COL3A1,
346 CXCL5, IL13, MMP7, SFTPA1, SFTPC, TIMP3).

347 **BAL Cell Type Imputation:** BAL contains an admixture of cell types in contact with the lumen of the
348 lower respiratory tract, and thus varying cell proportions may account for differential gene expression
349 detected by bulk sequencing. A multi-center study with small volume of BAL samples precluded single
350 cell sequencing. Hence, we used *in silico* cell type deconvolution with CIBERSORTx and the *Travaglini*
351 lung cell atlas to impute cell fractions in each sample.⁴²⁻⁴⁴ Consistent with findings described above, Cluster
352 1 showed high representation of antigen presenting cells including monocytes and macrophages, Cluster 2
353 showed a greater fraction of neutrophils, Cluster 3 showed a paucity of innate immune cells and a higher
354 fraction of CD4⁺ T-cells, and Cluster 4 showed a high fraction of CD8⁺ T-cells (**eFigure 11**). Given the
355 findings of varying cellular fraction within the BAL clusters, we then imputed cell-type specific gene
356 expression using CIBERSORTx (**Methods**). Monocyte-specific expression of the GOBP “Myeloid
357 Leukocyte Activation” gene set varied across clusters, with higher activation of activation markers such as
358 CSF1, IFNGR1, LDLR, TLR1, and TNF seen in Clusters 2, 3 and 4; notably, although Cluster 1 had a high
359 monocyte/macrophage cell fraction, lineage-specific inflammatory gene activation was relatively low in
360 this cluster (**eFigure 12**). Similarly, lymphocyte-specific expression of the GOBP “Lymphocyte
361 Activation” gene set varied across clusters, with the highest levels of markers such as AKT1, BTK, CD4,
362 DOCK8, JAK2, and IL7R seen in Clusters 3 and 4 (**eFigure 13**). Given the varying cell proportions and
363 imputed activation levels of lymphocytes across the clusters, we next aimed to determine whether there
364 might be differences in lymphocyte repertoires across the clusters. Using ImRep, we identified that the
365 majority of CDR3 alignments were for TCR α , with many fewer alignments to β , γ , and δ as well as to BCR
366 H, K, or L. Whereas the virally-enriched Cluster 4 showed the highest number of TRA clonotypes and
367 diversity, Cluster 1 showed the lowest (**eFigure 14**). Notably, BAL TCR $\alpha\beta$ clonotype numbers and
368 diversity were not correlated with blood lymphocyte count ($p=0.646$), although BAL TCR $\gamma\delta$ and BCR
369 subtypes were higher in patients with higher blood ALC ($p=0.041$ and $p=0.006$, respectively).

370 **Cluster transitions:** We next assessed whether original cluster assignments were stable over time. Thirty-
371 four patients had ≥ 2 BALs separated by a median of 79 days (IQR 21-243). Most patients who started in

372 the low-risk Cluster 1 moved out of Cluster 1 (17/26) to a higher-risk cluster, and patients who started
373 outside of Cluster 1 rarely moved into Cluster 1 (8/49), driving an overall change in the cluster burden over
374 time ($p < 0.001$, **eFigure 15, eTables 10-11**). This suggests that, for the subtype of patients with recurrent
375 or non-resolving lung disease, progression to an adverse BAL phenotype is common over time.

376 **Classification Model and External Cluster Validation:** Finally, as cluster assignments cannot be directly
377 applied to external cohorts, we used taxonomic and gene expression data to grow a random forest of 10,000
378 classification trees with a maximum depth of 10 nodes to be used as a cluster classifier. Out-of-bag AUC
379 was 0.923 indicating good cluster discrimination (**eTable 12**). Lung gene expression variables were
380 significantly more important to cluster classification than were taxonomic variables, with the 500 most
381 important genes showing significant enrichment for immune processes (**Data Files 12-13**). The random
382 forest classifier was then applied to taxonomic and gene expression data from an independent cohort of
383 $n=57$ BALs obtained from pediatric HCT recipients at the University Medical Center in Utrecht, the
384 Netherlands, between 2005-2016 (clinical traits described in **eTable 13**). Although this cohort differed in
385 geography, underlying diseases, allograft characteristics, and treatment protocols, 1-year non-relapse
386 mortality was lowest among patients with BALs assigned to the low-risk Cluster 1 (9%, 2/21), was higher
387 for patients assigned to the bacterially-rich Cluster 2 (36%, 4/11), and was highest for patients in the high-
388 risk Clusters 3 or 4 (52%, 13/25, $p=0.009$, **eFigure 16, eTable 14**), thus confirming the external validity
389 and clinical significance of the BAL cluster profiles.

390 **DISCUSSION**

391 Lung injury in pediatric hematopoietic cell transplant patients is frequently fatal, yet a lack of investigable
392 biospecimens has hindered progress in elucidating the pathobiology of disease. In this prospective
393 multicenter study, we used BAL from children at 32 hospitals to identify microbe-lung transcriptomic
394 signatures shared across patients. Although each BAL archetype was associated with undue morbidity,
395 microbial dysbiosis, undetected infection, and subtypes of inflammation and fibroproliferation were
396 identified as the primary hallmarks of fatal disease. Our findings come from a broad, international cohort

397 of children with poor immunity and high antimicrobial exposure and were replicated in an unrelated
398 validation cohort, thus lending credence to the work. These findings extend our previous work in pediatric
399 HCT candidates and suggest the possibility for precision pulmonary phenotyping as a key first step for
400 future interventional trials.

401 A major finding of our work is the identification of heterogeneous disease biology within a cohort of
402 medically complex patients where disease classification has been historically difficult.² BAL Cluster 1 was
403 most common, had moderate microbial burden, low rates of infection, predominantly alveolar macrophage-
404 related signaling, and the lowest mortality rates. In contrast, Cluster 2 showed high rates of microbial burden
405 and bacterial infections, higher blood neutrophil counts and BAL neutrophil-related gene expression, and
406 moderate mortality. Cluster 3 showed general microbiome depletion with enrichment of viruses and fungi
407 and fibroproliferative gene expression. Cluster 4 showed significant microbiome depletion with relative
408 sparing of Staphylococci and enrichment of viruses, commensurate with lymphocytic inflammation,
409 cellular injury, and the highest mortality rate. In the field of pulmonology, subclasses of asthma, acute
410 respiratory distress syndrome, and chronic obstructive pulmonary disease have recently been associated
411 with distinct clinical trajectories such that subclass-specific clinical trials are now emerging.⁴⁵⁻⁴⁷ The
412 identification of heterogeneous clusters may be the first step in improving bedside phenotyping and
413 ultimately enrolling pediatric HCT patients in biology-targeted interventional trials.

414 A second major finding of our work is the illumination of the delicate balance between the pulmonary
415 microbiome and mortality. The pulmonary microbiome is populated early in life by aerosolization of
416 oropharyngeal microbes during tidal ventilation, gastric aspiration, and disease-related hematogenous
417 spread.^{10,12,48,49} The near continuous exposure of the lungs to microbes introduces the opportunity for
418 infection but also supports immune and epithelial education in the form of tolerance and memory.^{50,51} The
419 ideal properties of the peri-HCT pulmonary microbiome likely require delicate balance between over-
420 population and eradication.^{10,12} Favoring the former, studies in cystic fibrosis and COPD have shown that
421 an increase in pulmonary microbial mass is associated with neutrophilic inflammation and disease

422 exacerbations⁵²⁻⁵⁵, a paradigm similar to patients in our bacterial- and neutrophil-enriched Cluster 2.
423 Favoring the latter, recent studies show that HCT patients with depleted or dysbiotic intestinal microbiomes
424 develop higher mortality rates due to excess colitis, graft versus host disease, and even pulmonary disease,
425 which is similar to patients in our Clusters 3 and 4.⁵⁶⁻⁵⁸ Our data show that a biodiversity and richness exist
426 reciprocally with pathogenic taxa such as *S.aureus*, *P.aeruginosa*, fungi, and viruses, suggesting that
427 commensal constituents may limit the ability for pathogens to expand,^{59,60} perhaps through local
428 immunomodulation or by direct nutrient competition.^{54,61-65} We show that the transcriptional activity of
429 BAL microbes is quite broad in patients with better clinical outcomes, raising the possibility that microbial
430 metabolites might benefit airway health, as has been recently shown for the anti-apoptotic microbial
431 metabolite indole-3-acetic acid (IAA).^{21,66,67}

432 Given the findings of commensal microbial depletion in non-survivors, we explored potential exposures
433 leading to this state. Antimicrobial exposure, the most likely culprit, has been strongly associated with
434 intestinal microbiome depletion and to a lesser extent pulmonary microbiome alterations mostly in the
435 cystic fibrosis and COPD populations.²⁸⁻³⁵ Our unsupervised analysis showed the highest antibiotic
436 exposure in the most bacterially-depleted BAL cluster 4, which complements the supervised analysis
437 finding of a negative relationship between AES and the quantity of numerous commensal bacteria. Causal
438 mediation modeling showed that association between AES and death was largely mediated by antibiotic-
439 induced contraction in BAL bacteria. Interestingly, we found that the quantity of the fungal phyla
440 Ascomycota increased with greater AES, supporting existing evidence that depletion of commensal
441 microbes may open a niche for opportunistic fungal growth.⁶⁸⁻⁷¹ Increased AES was associated with greater
442 BAL quantity of respiratory RNA viruses, consistent with previous associations between antibiotic
443 exposure and viral expansion.^{40,72} Although other factors such as the conditioning regimen may influence
444 microbiome composition,⁷³ these data argue for the need for judicious use of antibiotics, which might best
445 be achieved with rapid turnaround of clinical metagenomics assays in the future.⁷⁴ Certainly for critically

446 ill patients with unclear diagnoses, it will be difficult to feel confident in stopping antibiotics. Therefore,
447 microbiome-restorative therapies in patients necessarily antibiotic-exposed may be a crucial tool.⁷⁵

448 Over the past thirty years, numerous studies have confirmed that metagenomic sequencing for a wide range
449 of indications, such as meningitis and encephalitis, can increase diagnostic yield for pathogens.⁷⁶⁻⁷⁸

450 However, application of metagenomics to respiratory fluid has been hindered by difficulty discriminating
451 when a normal microbiome constituent such as *S.pneumoniae* expands to function as a pathogen. To address
452 this, we transformed our sequencing data from fractional to absolute using reference spike-ins and then
453 compared each microbe's detected level to that of other microbes in the sample (dominance) as well as to
454 other samples in the cohort (z-score). By parsing microbes in the context of the broader microbiome, we
455 provide a logical and intuitive approach to pathogen detection in non-sterile body sites. This approach
456 nearly doubled the number of patients with detected infections, while also providing a safeguard against
457 overcalling hits. Importantly, we identified novel viral strains, common and rare bacteria, and numerous
458 fungi and parasites as previously undetected causes of lung injury. Pathogen detection was highest in the
459 most dysbiotic clusters with the greatest commensal depletion and lowest richness and diversity, lending
460 credence to the above findings that airway commensals may safeguard the lungs against opportunistic
461 infections. Our data support the premise of a clinical trial using metagenomics to augment the clinical utility
462 of hospital diagnostics specifically in the setting of HCT. While many newly detected microbes have
463 existing effective treatments, many lack therapies at this time. Given our findings implicating antimicrobial
464 exposure with dysbiosis and poor clinical outcomes, antibiotic de-escalation or avoidance of dysbiosis may
465 be useful outcome metrics for such a trial.

466 The relationship between the pulmonary microbiome, lung epithelium, and the transplanted immune system
467 is characterized by a continuous and mutually influential interaction. In murine models of allogeneic HCT,
468 immune responses to pathogens can be both impaired as well as exaggerated, leading to delayed
469 phagocytosis, excessive myeloid cell recruitment and unremitting inflammation due to a lack of functional
470 NK- and T-cells.⁷⁹⁻⁸² Our data support this paradigm in human patients and reveal a complex and

471 heterogeneous immune response. Cluster 1, with a replete and diverse pulmonary microbiome, showed the
472 lowest mortality rates, low levels of granulocyte activation, and low levels of lymphocyte diversity and
473 lymphocyte-specific activation markers. In contrast, Cluster 2 showed neutrophil enrichment, and Clusters
474 3 and 4 showed a diverse lymphocyte population with markers of activation. Clinically, these distinctions
475 may be important, as patients might benefit from different approaches to immunomodulation. Notably,
476 Cluster 3 showed numerous markers of fibroproliferation and cellular senescence, suggesting transition to
477 a fibrotic phenotype that may merit treatment in upcoming clinical trials using novel anti-fibrotic agents.⁸³

478 This study has several limitations. First, the cohort's clinical heterogeneity requires interpreting findings
479 broadly. Second, clinical protocols were not standardized and thus varying post-HCT care across centers
480 could have influenced outcomes. Third, BAL collection was not standardized across centers and
481 bronchoscope controls were not obtained. However, our approach to adjusting for contamination used
482 ample internal and external controls. Fourth, detailed concurrent immunosuppressive regimens were not
483 collected. Fifth, despite methods to identify likely pulmonary pathogens, we could not adjudicate the
484 pathogenicity of each microbe or contribution to each patient's pulmonary disease. Sixth, clinical
485 microbiologic testing of BAL varied across hospitals and was not standardized. Seventh, while our work
486 implicates pulmonary microbial depletion in the pathobiology of post-HCT lung disease, we cannot prove
487 causality with correlative human studies, and cannot account for effects from other microbiomes such as
488 the intestinal microbiome on lung health.^{84,85}

489 In summary, we present the largest investigation to date of the pulmonary microbiome and transcriptome
490 in pediatric HCT patients. We identified four unique BAL clusters that combine microbiome and lung gene
491 expression signatures. The worst outcomes were observed for those with commensal microbe depletion,
492 viral or fungal enrichment, lymphocyte activation, and fibroproliferation. Overall, these findings represent
493 a step forward in understanding lung disease biology in HCT patients and may be used to improve patient
494 subtyping in preparation for a future biology-targeted clinical trial.

495

496 **METHODS**

497 **Patients:** The derivation cohort was enrolled through the Pediatric Transplantation and Cell Therapy
498 Consortium (PTCTC, NCT02926612) and the validation cohort was collected at the University Medical
499 Center in Utrecht, The Netherlands. Participating pediatric centers screened all patients with a history of
500 allogeneic (both cohorts) or autologous (PTCTC cohort only) HCT preparing to undergo clinically-
501 indicated bronchoscopic BAL for diagnostic assessment of pulmonary disease. Patients or their guardians
502 were approached prospectively for consent under local IRB approval at each site and permission was
503 obtained to collect leftover BAL fluid. Patients were excluded if there was a limitation of care such as do
504 not resuscitate at the time of BAL.

505 **BAL specimen collection:** Bronchoscopy and BAL were performed at the discretion of the treating team
506 using local institutional protocols. All BAL were obtained by pediatric pulmonologists trained in fiberoptic
507 bronchoscopy with anesthesia provided by anesthesiologists or critical care physicians. Lavage protocol
508 was not dictated by the study but typically involved 3-6 aliquots of 10mL sterile saline inserted into diseased
509 areas of the lung as determined by preceding chest imaging or physical exam.⁸⁶ Percent of lavage returned
510 was not routinely documented and lavage aliquots were typically pooled by the clinical team immediately
511 after collection.^{87,88} After aliquoting for clinical testing, excess lavage was placed immediately on dry ice,
512 stored at -70°, shipped to UCSF, and stored at -70°C until processing.

513 **Clinical protocols and data collection:** Clinical microbiologic testing was determined by the treating team
514 and typically included culture for bacteria, fungus, and AFB; multiplex PCR for respiratory viruses;
515 galactomannan antigen; and cytology for PCP. Additional molecular diagnostics such as PCR for atypical
516 bacteria or fungi were used at the discretion of the site. After BAL, supportive care protocols were
517 determined by the treating team; all patients were enrolled at centers with pediatric intensive care units.
518 Patient demographics, medical history, and transplant-specific data were documented by trained study
519 coordinators at each site. The most recent ANC and ALC measured clinically prior to BAL were
520 documented. Results of clinical microbiologic testing on BAL were documented and not considered

521 complete until 4 weeks after collection. For the PTCTC cohort, all doses of antimicrobials administered in
522 the 7 days prior to BAL were documented. Antibiotic exposure score (AES) was calculated by summing
523 days of exposure to each antibacterial agent weighted with an agent-specific broadness score ranging from
524 4 to 49.75 (e.g.: ampicillin 13.50, meropenem 41.50).³⁸ Daily dosages were not collected. The number of
525 anti-anaerobe days were calculated as the sum of preceding exposure to each of the following:
526 Amoxicillin/clavulanic acid, Ampicillin/sulbactam, Piperacillin/tazobactam, Meropenem, Ertapenem,
527 Imipenem, Levofloxacin, Clindamycin, Doxycycline, Tigecycline, or Metronidazole. Patients were
528 followed until hospital discharge (PTCTC) or until at least one year post-BAL (Utrecht) with no loss to
529 follow-up.

530 **BAL RNA Extraction:** Samples were used on the first or second thaw. Samples underwent a previously
531 described RNA extraction protocol optimized for BAL fluid.¹¹ 200 μ L of BAL was combined with 200 μ L
532 DNA/RNA Shield (Zymo) and 0.5mm glass bashing beads (Omni) for 5 cycles of 25 seconds bashing at
533 30Hz, with 60 seconds of rest on ice between each cycle (TissueLyser II, Qiagen). Subsequently, samples
534 were centrifuged for 10 minutes at 4°C and the supernatant was used for column-based RNA extraction
535 with DNase treatment according to the manufacturer's recommendations (Zymo ZR-Duet DNA/RNA
536 MiniPrep Kit). Resultant RNA was eluted in 5 μ L sterile water and stored at -70°C until sequencing library
537 preparation.

538 **BAL RNA Sequencing:** Samples underwent a previously described sequencing library preparation
539 protocol optimized for BAL fluid.¹⁹ First, BAL RNA was dehydrated at 40°C for 25 minutes in a 384 well
540 plate (GeneVac E-Z2). Second, sequencing libraries were prepared using miniaturized protocols adapted
541 from the New England Biolabs Ultra II RNA Library Prep Kit (dx.doi.org/10.17504/protocols.io.tcaeise).
542 Reagents were dispensed using the Echo 525 (Labcyte) and underwent Ampure-XP bead cleaning on a
543 Beckman Coulter Biomek NX^P instrument. Libraries underwent 19 cycles of polymerase chain reaction
544 (PCR) amplification, size selection to a target 300 to 700 nucleotides (nt), and were pooled to facilitate
545 approximately even depth of sequencing. Twenty-five picograms (pg) of External RNA Controls

546 Consortium (ERCC) pooled standards were spiked-in to each sample after RNA extraction and before
547 library preparation to serve as internal positive controls (Thermo Fisher Scientific Cat. No 4456740). In
548 addition, to identify contamination in laboratory reagents and the laboratory environment, each batch
549 contained 2 samples of 200 μ L sterile water and 6-8 samples of 200 μ L HeLa cells taken from a laboratory
550 stock and processed identically to patient samples, in order to account for laboratory- and reagent-
551 introduced contamination. These samples were processed at the same time as the patient BAL samples in
552 order to use the same lot of reagents and minimize batch effect on control samples. Samples were processed
553 and sequenced in 4 batches. Samples were pooled across lanes of an Illumina NovaSeq 6000 instrument
554 and sequenced to a target depth of 40 million read-pairs with sequencing read length of 125 nt.

555 **Sequencing file processing**

556 All sequencing files were processed using the CZID pipeline v7.1 ([https://github.com/chanzuckerberg/czid-](https://github.com/chanzuckerberg/czid-web)
557 [web](https://github.com/chanzuckerberg/czid-web)).¹⁸ Briefly, .fastq files underwent a first round of human read subtraction (*STAR* to hg38) followed by
558 Illumina adaptor removal (*Trimmomatic*), quality filtering (*PriceSeq*), duplicate read removal (*CD-HIT-*
559 *DUP*), and LZW complexity filtering. Next, sequencing files underwent a second more stringent round of
560 human read subtraction (*Bowtie2*) followed by a third round of human read subtraction (*STAR*),
561 subsampling to 1 million fragments, and a fourth and final round of human read subtraction (*GSNAP*).
562 Human gene counts were produced using the CZID pipeline with alignment to hg38 as described above.
563 60,590 total genes were detected across all samples (median 44,063, IQR 31,553-52,129), and were subset
564 for 19,032 protein coding genes (median 18,259 genes per sample, IQR 16,988-18,871) and used in
565 analyses below. Resultant human-subtracted sequencing files were then used in two ways for microbiome
566 characterization:

567 **(1) Microbial taxonomic alignment:** Human-subtracted sequencing files underwent alignment to the NCBI
568 nt/nr database using *GSNAP* with minimum alignment length >36. Quality metrics for the sequencing run
569 including percent of reads that passed the *PriceSeq* filter step and percent of reads that passed all steps were
570 examined and samples with poor sequencing quality were re-sequenced. Taxa counts were generated with

571 associated metrics of percent identity, contig length, and e-value to the nearest NCBI hit. To reduce spurious
572 associations due to ambiguous alignments, taxa were excluded if they (1) aligned to archaea or uncultured
573 microbes, (2) had ≤ 6 total reads, (3) had < 100 nt alignment length, or (4) had $< 80\%$, $< 90\%$, or $< 95\%$ nt
574 percent identity for viruses, eukaryotes, and bacteria, respectively. In addition, samples with low biomass
575 (< 100 pg) were further filtered to keep only taxa with ≥ 10 transcripts forming a contig of ≥ 250 nt with $\geq 80\%$
576 percent identity to the nearest NCBI hit.

577 **(2) Microbial functional alignment:** Human-subtracted sequencing files were processed using FMAP
578 v.0.15⁸⁹ in order to profile the metabolic pathways present in each sample. FMAP_mapping.pl was paired
579 with diamond v.0.9.24⁹⁰ and FMAP_quantification.pl were used with default settings to identify and
580 quantify associated proteins in the UniRef90 database.^{91,92} The gene assignments were regrouped by KEGG
581 descriptors^{93,94} and their annotation was summarized at levels 1 to 3.

582 **Microbial quantification and contamination**

583 Low biomass samples are susceptible to contamination.^{25,27} We previously showed that a positive control
584 spike-in to each sample can be used to back-calculate the original RNA mass of the sample by solving the
585 linear proportionality equation (total sample reads / total sample mass) \approx (ERCC reads / ERCC mass),
586 where sample reads and ERCC reads were detected by the above protocol and ERCC input was standardized
587 as 25 pg.²⁰ The calculated sample mass was then reduced by 25 pg (the ERCC input) to equal the original
588 sample mass before ERCC addition. Since the input RNA mass of the water controls was determined to be
589 about (5 pg presumably reflecting 5 pg of sequenceable contamination), we discarded samples whose total
590 input mass was below 10 pg, as we were unable to reliably differentiate between contamination and true
591 constituents. Since low biomass samples will preferentially amplify contaminants, we then used the ERCC
592 spike-in to transform reads into estimated mass, allowing analysis of both fractional and absolute
593 microbiome properties. Since each BAL microbiome consists of contributions from the patient and
594 externally introduced contaminants, we then calculated the unique contamination profile of the water and

595 HeLa samples for each sequencing batch, and subtracted the mean + 2SD of each contaminant taxa from
596 the patient samples processed in the respective batch. Mass-transformed and contamination-adjusted values
597 were used for downstream analysis.

598 **Statistical Analysis**

599 **(1) Unsupervised Clustering Analysis:** Since microbiome data can be described using taxonomy,
600 functional annotation, or summary measures, we used the Multi-Omics Factor Analysis to reduce
601 dimensionality and identify a core set of factors.⁹⁵ This approach accommodates different data structures
602 and distributions and is tolerant of collinearity. Data were filtered to include phyla, genera, species, and
603 KEGG pathways present in >15% of samples, underwent variance stabilizing transformation (*vst*, *DESeq2*),
604 and were combined with aggregate metrics of total microbial mass, Simpson's and Shannon's alpha
605 diversity (*vegan*), and richness, which was defined as number of species detected at a threshold of ≥ 1
606 pg.^{96,97} MOFA was used to identify 15 core latent factors that together explained the most variance in the
607 data structure. The matrix of latent factor values then underwent uniform manifold approximation mapping
608 (*umap*) and BAL clusters were identified using hierarchical clustering of euclidean distances (*eclust*,
609 *factoextra*). The ideal number of clusters was determined to be four using the silhouette, elbow, and gap-
610 statistic plots.

611 **(2) Clinical characteristics:** Kaplan Meier survival analysis was used to plot in-hospital mortality by BAL
612 cluster and survival curves were compared using the log-rank test of equality (*survival*). Differences in
613 clinical traits across clusters (eg: antimicrobial exposure score, absolute neutrophil count) were tested using
614 the non-parametric Kruskal-Wallis (*kruskaltests*) and Dunn's tests (*dunn.test*) or Chi-squared test as
615 appropriate.

616 **(3) Microbiome comparisons:** Differences in microbial taxa, KEGG pathways, richness, and diversity
617 across the 4 BAL clusters were tested using the non-parametric Kruskal-Wallis (*kruskaltests*) and Dunn's
618 tests (*dunn.test*) with Benjamini-Hochberg correction for multiple hypothesis testing. Differences in

619 microbial taxa and KEGG pathways were also tested using negative binomial generalized linear models,
620 which account for both microbiome composition and size by inclusion of taxa-specific dispersion factors
621 (*edgeR*).⁹⁸ Associations between microbial taxa and clinical variables (e.g.: antimicrobial exposure score,
622 in-hospital mortality) were tested using *edgeR*. Data were visualized with heatmaps showing cluster means
623 for each variable (*pheatmap*) with individual comparisons shown using box-whisker plots (*ggplot*). Causal
624 mediation was used to test whether the association between antimicrobial exposure and mortality was
625 mediated by an antibiotic-induced reduction in certain BAL microbes (*mediation*).^{99,100} Using the latent
626 structural equation framework, we fit (1) poisson models for the association between preceding AES and
627 BAL quantity of a certain microbe, and (2) logistic regression models for the association between BAL
628 quantity of a given microbe and outcome, independent of AES. Mediation was tested using 1,000
629 simulations with bootstrapped confidence intervals and direct and indirect effects were plotted (**eFigure 7**).

630 **(4) Pathogen identification:** Taxa considered as potential respiratory pathogens were adapted from the
631 CZID Pathogen List (https://czid.org/pathogen_list) with modifications for immunocompromised patients
632 and pathogens specific to the respiratory system. The final list of taxa considered is detailed in **Data File**
633 **4**. We did not include avirulent viruses, such as TTV, or bacterial commensals that are infrequently a cause
634 of pulmonary disease, such as *Prevotella* species, coagulase-negative *Staphylococci*, non-diphtheria
635 *Corynebacterium*, and viridans group *Streptococci*, although these have at times been implicated in
636 pulmonary disease in immunocompromised patients. To identify potentially pathogenic viruses, we applied
637 a threshold of viral detection at any level above background (after applying the quality and contamination
638 filters described above). This presence/absence approach was selected to mirror the approach used in
639 clinical respiratory viral panels, which typically dichotomizes any level of detection as present/absent. To
640 identify potentially pathogenic bacteria, we applied a threshold of detection with mass $\geq 10\text{pg}$, bacterial
641 dominance $\geq 20\%$, and Z-score $\geq +2$, where Z-score was calculated as the number of standard deviations
642 above the mean of the \log_{10} -transformed mass values for each microbe in the cohort. Requiring a minimum
643 mass, dominance, and z-score was based on the historical framework that bacterial infections occur when

644 microbes are present at high mass that is greater than other microbes and greater than in other (non-infected)
645 patients, although this may not be true in all instances. Cutoff values were empirically selected after analysis
646 of data distributions and could be exchanged for other cutoffs in order to alter the balance between
647 sensitivity and specificity of calls. Finally, to identify potentially pathogenic fungi, we applied a threshold
648 of detection with mass $\geq 10\text{pg}$ and Z-score $\geq +2$. We did not apply a microbiome dominance cutoff for
649 fungal pathogens since the relationship between organisms in the pulmonary mycobiome is less well
650 understood.

651 **(5) Gene expression:** Only genes present in $>25\%$ of samples were used for differential gene expression.
652 To identify individual differentially expressed genes, we used a 4-way ANOVA-like approach with
653 negative binomial generalized linear models (*edgeR*). Select differentially expressed genes identified at a
654 threshold FDR ≤ 0.05 were visualized with box-whisker plots of variance stabilization-transformed counts.
655 To compute gene set enrichment scores, we used non-parametric gene set variation analysis with Poisson
656 distributions (*gsva*) and the Reactome set of $n=1,554$ gene sets.^{101,102} Differences in enrichment scores
657 across the BAL clusters were compared using Kruskal-Wallis (*kruskaltests*) and Dunn's tests (*dunn.test*)
658 and gene sets with significant differences were visualized using dot plots of the mean expression scores
659 (*pheatmap*). Next, cell types contributing to bulk seq expression were imputed using CIBERSORTx
660 (Docker version), which employs a user-defined reference single-cell atlas to identify cell-type specific
661 transcript ratios and impute cell fractions (we selected the *Travaglini et al* lung cell atlas).⁴²⁻⁴⁴ Cell-type
662 specific gene expression was imputed using CIBERSORTx high resolution mode, which utilizes previously
663 created cell fractions to impute cell-type specific expression. Finally, lymphocyte receptor repertoires were
664 imputed using Imrep (Linux install), which identifies CDR3 alignments from within bulk gene expression
665 data.^{103,104}

666 **(6) Classification and validation:** Since cluster assignments cannot be directly applied to an external
667 dataset, a classification tool is required to predict cluster assignments. We trained a random forest of
668 $n=10,000$ trees using microbiome taxonomy and lung gene expression datasets as input, and 2x weighting

669 of clusters 3 and 4 given the BAL cluster imbalance (*randomforestSRC*).^{105,106} Ideal forest parameters
670 determined using *tune* were similar to default settings and thus default settings were used for all other
671 parameters (eg: *mtry*, *nodesize*, etc). Forest accuracy was determined using out of bag AUC and a confusion
672 matrix. Variable importance was determined using permutation VIMP (Breiman-Cutler importance) by
673 permuting OOB cases (*vimp*). To validate the classifier, the random forest classifier was applied to
674 microbiome taxonomy and lung gene expression data from the n=57 Utrecht BALs and 1-year post-BAL
675 non-relapse mortality rates were compared according to predicted BAL cluster type using Kaplan-Meier
676 survival curves with the log-rank test.

677 **DATA AVAILABILITY:** Sequencing files are posted on dbGaP:

678 https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001684.v2.p1

679 **Table 1. Patient Characteristics**

Demographics (n=229 patients)	
Age (median years, IQR)	11.0 (IQR 4.7-16.7)
Sex (male)	133 (58.15%)
Race	
- White	140 (61.1%)
- Black	29 (12.7%)
- Other/multiple	26 (11.4%)
- Asian/PI	25 (10.9%)
- Native American	2 (0.9%)
- Unknown	7 (3.1%)
Ethnicity - Latino/Hispanic	59 (25.8%)
Medical History (n=229 patients)	
Disease	
- Leukemia ^a	125 (54.6%)
- Inborn errors of immunity ^b	40 (17.5%)
- Non-malignant hematologic ^c	27 (11.8%)
- Solid tumor ^d	14 (6.1%)
- Lymphoma ^e	12 (5.2%)
- Inborn errors of metabolism ^f	11 (4.8%)
HCT Type	
- Allogeneic	213 (93.0%)
- Bone marrow	- 92 (43.2%)
- Peripheral blood	- 88 (41.3%)
- Umbilical cord blood (UCB)	- 33 (15.5%)
- Autologous	16 (7.0%)
HLA match (allogeneic only)	
- Matched related donor	45 (21.1%)
- Matched unrelated donor (inc. 6/6 UCB)	49 (23.0%)
- Mismatched related donor (haplo)	57 (26.8%)
- Mismatched unrelated donor (inc. <6/6 UCB)	61 (29.1%)
Conditioning Agents Used ^g	
- Backbone agent	
- Busulfan	86 (37.6%)
- Melphalan	146 (63.8%)
- Total body irradiation (TBI)	63 (27.5%)
- Other ^h	20 (8.7%)
- Other alkylating agent	
- Cyclophosphamide	91 (39.7%)
- Thiotepa	66 (28.8%)
- Antimetabolite	
- Clofarabine	15 (6.6%)
- Cytarabine	5 (2.2%)
- Fludarabine	146 (63.8%)
- Serotherapy (ATG or Alemtuzumab)	119 (52.0%)

680

681 **Legend:** ^a includes B-ALL (n=54), AML (n=39), T-ALL (n=12), JMML (n=6), CML (n=4),

682 other/MDS/mixed phenotype (n=10). ^b includes SCID (n=14), HLH (n=7), CGD (n=4), WAS (n=3), other

683 (n=12). ^c includes SAA (n=12), Fanconi anemia (n=4), sickle cell disease (n=9), thalassemia (n=2). ^d
684 includes neuroblastoma (n=10), medulloblastoma (n=3), other solid tumor (n=1). ^e includes B-cell
685 lymphoma (n=6), non-EBV T-cell lymphoma (n=4), EBV+ T-cell lymphoma (n=2). ^f includes Hurler
686 syndrome (n=4), osteopetrosis (n=2), X-linked adrenoleukodystrophy (n=2), other (n=3). ^g Patients may
687 have received multiple agents in the same or multiple categories. ^h includes carmustine (n=2), treosulfan
688 (n=3), carboplatin (n=4), etoposide (n=16).

689 **Table 2. Clinical Presentation and Outcomes**

Characteristics at Enrollment (n=278 events with BAL)	
Days from HCT to BAL (n, %)	114 (IQR 36-331)
Days from Symptoms to BAL ^a (n, %)	8 (IQR 2-21)
Clinical Presentation Symptoms (n, %)	
- Lower respiratory symptoms (cough, tachypnea, etc) ^b	249 (89.7%)
- Hypoxia \leq 96%	202 (72.7%)
- Abnormal chest x-ray ^c	174/207 (84.1%)
- Abnormal chest CT ^d	209/218 (95.9%)
- Worsening PFTs	16 (5.8%)
Respiratory Support Prior to BAL (n, %)	
- No oxygen	159 (56%)
- Nasal cannula or face mask	41 (14%)
- High-flow nasal cannula	19 (7%)
- Non-invasive positive pressure (CPAP or BiPAP)	11 (4%)
- Endotracheal intubation with mechanical ventilation	54 (19%)
Comorbidities at time of BAL (n, %)	
- Engraftment syndrome	15 (5.4%)
- GVHD active at time of BAL ^e	83/260 (31%)
- GVHD ever preceding BAL	126/260 (48.5%)
- Heart failure or reduced function	11 (4.0%)
- Kidney injury	47 (16.9%)
- Pericardial effusion	25 (9.0%)
- Pulmonary hemorrhage/hemoptysis	23 (8.3%)
- Sepsis	37 (13.3%)
- TA-TMA	22 (7.9%)
- VOD/SOS	24 (8.6%)
Immunologic Function Prior to BAL ^f	
- WBC (median, IQR)	4,415 (2,370-8,400)
- ANC (median, IQR)	3,060 (1,632-5,508)
- ANC $<0.5 \times 10^9/L$ (n, %)	34 (12.2%)
- ALC (median, IQR)	420 (156-1,035)
- ALC $<0.2 \times 10^9/L$ (n, %)	77 (27.7%)
BAL Clinical Microbiology Results (n, %)	
- Any positive	116 (41.7%)
- Bacterial	51 (18.3%)
- Viral	76 (27.3%)
- Fungal/Protozoal	25 (9.0%)
- More than 1 organism	29 (10.4%)
Antimicrobials in Preceding 1 Week (median, IQR)	
- Antibacterial	3 (IQR 2-4, range 0-9)
- Antivirals	1 (IQR 1-2, range 0-3)
- Antifungals	1 (IQR 0-1, range 0-3)
Outcomes (n=229 patients)	
Required intensive care	121 (52.8%)
Required \geq 7 days mechanical ventilation	71 (31.0%)
In-hospital mortality	45 (19.7%)

691 **Legend:** ^a missing in n=14. ^b n=29 patients without clinical symptoms underwent BAL to evaluate declining
692 PFTs or chest CT abnormalities. ^{c,d} chest-xray and chest CT obtained prior to n=207 and n=218 BALs,
693 respectively. ^e GVHD assessed in allograft recipients only. ^f WBC, ANC, ALC expressed as 10⁹ cells/L
694 whole blood.

695 **FIGURE LEGENDS**

696 **Figure 1. Study design and clinical outcomes.** (A) Patients were recruited from 32 participating children's
697 hospitals in the United States, Canada, and Australia. (B) Study design concept diagram. (C) BAL
698 processing and analysis workflow. (D) Four microbiome-transcriptome clusters were identified. (E) In-
699 hospital survival for all patients (left) and the subset requiring respiratory support prior to testing (right)
700 was plotted according to BAL cluster and differences were analyzed with the log rank test.

701 **Figure 2. BAL microbiome.** (A) The fraction (left) and mass (right) of major bacterial, viral, and fungal
702 phyla are plotted, with shading representing the average for each of the 4 BAL clusters. The average mass
703 of bacterial genera in each of the 4 BAL clusters are shown to the right. (B) Taxonomic richness and
704 diversity are plotted across the 4 BAL clusters. (C) Microbes associated with in-hospital mortality were
705 identified using negative binomial generalized linear models (*edgeR*) and are plotted according to logFC
706 (position, color) and FDR (dot size). (D) Taxonomic richness and diversity stratified by survival status. (E)
707 Microbial alignments to KEGG metabolic pathways were averaged for each BAL cluster. (F) Select
708 metabolic pathways that differ across the BAL clusters are shown.

709 **Figure 3. BAL pathogen detection.** (A) Left: Dotplots of common community-transmitted respiratory
710 viruses (left), herpesviruses (middle), and all other viruses (right) detected in the cohort, plotted according
711 to microbial mass (x-axis) and microbiome dominance (y-axis). Right: A bar chart comparing viral
712 detection across the 4 BAL clusters according to hospital tests and metagenomic sequencing. (B) Left: All
713 *H.influenzae*, *S.aureus*, and *S.pneumoniae* detected in the cohort are plotted, with dotted lines indicating
714 cutoffs of mass $\geq 10\text{pg}$ and bacterial dominance $\geq 20\%$. Taxa above these cutoffs are shown in the upper-
715 right quadrant (shaded yellow) to indicate outliers within the cohort. Right: A bar chart comparing
716 potentially pathogenic bacteria detected across the 4 BAL clusters according to hospital tests and
717 metagenomic sequencing. (C) Left: All microbes detected in BAL of three patients are shown, with arrows
718 pointing to fungi present in high quantities. Right: A bar chart comparing potentially pathogenic eukaryotes
719 detected across the 4 BAL clusters according to hospital tests and metagenomic sequencing.

720 **Figure 4. Antibiotic exposure and impact on BAL microbiome.** (A) Days of antimicrobials are listed for
721 antibacterials (black), antifungals (green), and antivirals (blue). Patients are listed in columns and shading
722 indicates number of days of exposure to each antibiotic in the week preceding BAL. (B) Antibiotic exposure
723 score (AES) was calculated prior to each BAL as the sum of antibiotic exposure days times a broadness
724 weighting factor, summed for all therapies received in the week preceding BAL. AES varied across the
725 clusters and was highest for patients in Cluster 4. (C) Negative binomial generalized linear models were
726 used to test for BAL microbes associated with AES. Microbes are listed in rows, with phyla shown on the
727 left and bacterial genera shown on the right.

728 **Figure 5. BAL gene expression.** (A) Differentially expressed genes were identified by 4-way ANOVA
729 like analysis with negative binomial generalized linear models. Mean normalized expression levels for
730 significant genes are displayed for the 4 BAL clusters. (B) Individual differentially expressed genes were
731 identified across the 4 clusters (*edgeR*) and variance-stabilized transformed gene counts for select genes
732 highest in each of the 4 clusters are plotted. (C) Gene set enrichment scores to Reactome pathways were
733 calculated and example gene sets most enriched in each of the 4 clusters are shown.

734

735 **REFERENCES**

- 736 1. Jenq RR, van den Brink MRM. Allogeneic haematopoietic stem cell transplantation: individualized
737 stem cell and immune therapy of cancer. *Nat Rev Cancer*. 2010 Mar;10(3):213–221. PMID:
738 20168320
- 739 2. Panoskaltis-Mortari A, Griese M, Madtes DK, Belperio JA, Haddad IY, Folz RJ, Cooke KR,
740 American Thoracic Society Committee on Idiopathic Pneumonia Syndrome. An official American
741 Thoracic Society research statement: noninfectious lung injury after hematopoietic stem cell
742 transplantation: idiopathic pneumonia syndrome. *Am J Respir Crit Care Med*. United States;
743 2011;183(9):1262–1279.
- 744 3. Fitch T, Myers KC, Dewan M, Towe C, Dandoy C. Pulmonary Complications After Pediatric Stem
745 Cell Transplant. *Front Oncol*. 2021;11:755878. PMID: PMC8550452
- 746 4. Walker H, Shanthikumar S, Cole T, Neeland M, Hanna D, Haeusler GM. Novel approaches to the
747 prediction and diagnosis of pulmonary complications in the paediatric haematopoietic stem cell
748 transplant patient. *Curr Opin Infect Dis*. 2022 Dec 1;35(6):493–499. PMID: 36345851
- 749 5. Broglie L, Fretham C, Al-Seraihy A, George B, Kurtzberg J, Loren A, MacMillan M, Martinez C,
750 Davies SM, Pasquini MC. Pulmonary Complications in Pediatric and Adolescent Patients Following
751 Allogeneic Hematopoietic Cell Transplantation. *Biol Blood Marrow Transplant J Am Soc Blood*
752 *Marrow Transplant*. 2019 Oct;25(10):2024–2030. PMID: PMC7262781
- 753 6. Kaya Z, Weiner DJ, Yilmaz D, Rowan J, Goyal RK. Lung function, pulmonary complications, and
754 mortality after allogeneic blood and marrow transplantation in children. *Biol Blood Marrow*
755 *Transplant J Am Soc Blood Marrow Transplant*. United States; 2009;15(7):817–826.

- 756 7. Zinter MS, Logan BR, Fretham C, Sapru A, Abraham A, Aljurf MD, Arnold SD, Artz A, Auletta
757 JJ, Chhabra S, Copelan E, Duncan C, Gale RP, Guinan E, Hematti P, Keating AK, Marks DI, Olsson
758 R, Savani BN, Ustun C, Williams KM, Pasquini MC, Dvorak CC. Comprehensive Prognostication
759 in Critically Ill Pediatric Hematopoietic Cell Transplant Patients: Results from Merging the Center
760 for International Blood and Marrow Transplant Research (CIBMTR) and Virtual Pediatric Systems
761 (VPS) Registries. *Biol Blood Marrow Transplant J Am Soc Blood Marrow Transplant. United*
762 *States*: . Published by Elsevier Inc; 2020;26(2):333–342.
- 763 8. Zinter MS, Brazauskas R, Strom J, Chen S, Bo-Subait S, Sharma A, Beitinjaneh A, Dimitrova D,
764 Guilcher G, Preussler J, Myers K, Bhatt NS, Ringden O, Hematti P, Hayashi RJ, Patel S, De Oliveira
765 SN, Rotz S, Badawy SM, Nishihori T, Buchbinder D, Hamilton B, Savani B, Schoemans H, Sorrow
766 M, Winestone L, Duncan C, Phelan R, Dvorak CC. Critical Illness Risk and Long-Term Outcomes
767 Following Intensive Care in Pediatric Hematopoietic Cell Transplant Recipients. *MedRxiv Prepr*
768 *Serv Health Sci*. 2023 Aug 5;2023.07.31.23293444. PMID: PMC10418579
- 769 9. Khemani RG, Smith L, Lopez-Fernandez YM, Kwok J, Morzov R, Klein MJ, Yehya N, Willson D,
770 Kneyber MCJ, Lillie J, Fernandez A, Newth CJL, Jouvet P, Thomas NJ, Pediatric Acute Respiratory
771 Distress syndrome Incidence and Epidemiology (PARDIE) Investigators, Pediatric Acute Lung
772 Injury and Sepsis Investigators (PALISI) Network. Paediatric acute respiratory distress syndrome
773 incidence and epidemiology (PARDIE): an international, observational study. *Lancet Respir Med*.
774 2019 Feb;7(2):115–128. PMID: PMC7045907
- 775 10. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The Microbiome and the Respiratory
776 Tract. *Annu Rev Physiol. United States*; 2016;78:481–504.
- 777 11. Zinter MS, Dvorak CC, Mayday MY, Iwanaga K, Ly NP, McGarry ME, Church GD, Faricy LE,
778 Rowan CM, Hume JR, Steiner ME, Crawford ED, Langelier C, Kalantar K, Chow ED, Miller S,
779 Shimano K, Melton A, Yanik GA, Sapru A, DeRisi JL. Pulmonary Metagenomic Sequencing

- 780 Suggests Missed Infections in Immunocompromised Children. *Clin Infect Dis Off Publ Infect Dis*
781 *Soc Am. United States*; 2019;68(11):1847–1855.
- 782 12. Natalini JG, Singh S, Segal LN. The dynamic lung microbiome in health and disease. *Nat Rev*
783 *Microbiol.* 2023 Apr;21(4):222–235. PMID: PMC9668228
- 784 13. Kelly MS, Spees L, Vinesett R, Stokhuyzen A, McGill L, Proia AD, Jenkins K, Arshad M, Seed PC,
785 Martin PL. Utility of Autopsy among Pediatric Allogeneic Hematopoietic Stem Cell Transplant
786 Recipients: One Last Chance to Learn? *Biol Blood Marrow Transplant J Am Soc Blood Marrow*
787 *Transplant.* 2018 Sep;24(9):1861–1865. PMID: PMC6163060
- 788 14. Multani A, Allard LS, Wangjam T, Sica RA, Epstein DJ, Rezvani AR, Ho DY. Missed diagnosis
789 and misdiagnosis of infectious diseases in hematopoietic cell transplant recipients: an autopsy study.
790 *Blood Adv.* 2019 Nov 26;3(22):3602–3612. PMID: PMC6880905
- 791 15. Langelier C, Zinter MS, Kalantar K, Yanik GA, Christenson S, O’Donovan B, White C, Wilson M,
792 Sapru A, Dvorak CC, Miller S, Chiu CY, DeRisi JL. Metagenomic Sequencing Detects Respiratory
793 Pathogens in Hematopoietic Cellular Transplant Patients. *Am J Respir Crit Care Med. United States*;
794 2018;197(4):524–528.
- 795 16. Zinter MS, Lindemans CA, Versluys BA, Mayday MY, Sunshine S, Reyes G, Sirota M, Sapru A,
796 Matthay MA, Kharbanda S, Dvorak CC, Boelens JJ, DeRisi JL. The pulmonary metatranscriptome
797 prior to pediatric HCT identifies post-HCT lung injury. *Blood.* 2021 Mar 25;137(12):1679–1689.
798 PMCID: PMC7995292
- 799 17. Zinter MS, Versluys AB, Lindemans CA, Mayday MY, Reyes G, Sunshine S, Chan M, Fiorino EK,
800 Cancio M, Prevaes S, Sirota M, Matthay MA, Kharbanda S, Dvorak CC, Boelens JJ, DeRisi JL.
801 Pulmonary microbiome and gene expression signatures differentiate lung function in pediatric

- 802 hematopoietic cell transplant candidates. *Sci Transl Med*. 2022 Mar 9;14(635):eabm8646. PMID:
803 35263147
- 804 18. Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, Han J, Holmes OB, Juan
805 YF, King R, Kislyuk A, Lin MF, Mariano M, Morse T, Reynoso LV, Cruz DR, Sheu J, Tang J,
806 Wang J, Zhang MA, Zhong E, Ahyong V, Lay S, Chea S, Bohl JA, Manning JE, Tato CM, DeRisi
807 JL. IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen
808 detection and monitoring. *GigaScience*. 2020 15;9(10):giaa111. PMCID: PMC7566497
- 809 19. Mayday MY, Khan LM, Chow ED, Zinter MS, DeRisi JL. Miniaturization and optimization of 384-
810 well compatible RNA sequencing library preparation. *PloS One*. United States;
811 2019;14(1):e0206194.
- 812 20. Zinter MS, Mayday MY, Ryckman KK, Jelliffe-Pawlowski LL, DeRisi JL. Towards precision
813 quantification of contamination in metagenomic sequencing experiments. *Microbiome*. England;
814 2019;7(1):62–6.
- 815 21. Sulaiman I, Wu BG, Li Y, Tsay JC, Sauthoff M, Scott AS, Ji K, Koralov SB, Weiden M, Clemente
816 JC, Jones D, Huang YJ, Stringer KA, Zhang L, Geber A, Banakis S, Tipton L, Ghedin E, Segal LN.
817 Functional lower airways genomic profiling of the microbiome to capture active microbial
818 metabolism. *Eur Respir J*. 2021 Jul;58(1):2003434. PMCID: PMC8643072
- 819 22. Rhinovirus C13 strain RvC13/USA/2019, complete genome [Internet]. 2023 [cited 2023 Oct 3].
820 Available from: <http://www.ncbi.nlm.nih.gov/nuccore/OQ116581.1>
- 821 23. Rhinovirus C17 strain RvC17/USA/2019, complete genome [Internet]. 2023 [cited 2023 Oct 3].
822 Available from: <http://www.ncbi.nlm.nih.gov/nuccore/OQ116583.1>

- 823 24. Rhinovirus A56 strain RvA56/USA/2019, complete genome [Internet]. 2023 [cited 2023 Oct 3].
824 Available from: <http://www.ncbi.nlm.nih.gov/nucore/OQ116582.1>
- 825 25. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low
826 Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol.* England:
827 Elsevier Ltd; 2019;27(2):105–117.
- 828 26. Weiss S, Amir A, Hyde ER, Metcalf JL, Song SJ, Knight R. Tracking down the sources of
829 experimental contamination in microbiome studies. *Genome Biol.* England; 2014;15(12):564–2.
- 830 27. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and
831 removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome.* England;
832 2018;6(1):226–2.
- 833 28. Flanagan JL, Brodie EL, Weng L, Lynch SV, Garcia O, Brown R, Hugenholtz P, DeSantis TZ,
834 Andersen GL, Wiener-Kronish JP, Bristow J. Loss of bacterial diversity during antibiotic treatment
835 of intubated patients colonized with *Pseudomonas aeruginosa*. *J Clin Microbiol.* 2007
836 Jun;45(6):1954–1962. PMID: PMC1933106
- 837 29. Hernández-Terán A, Vega-Sánchez AE, Mejía-Nepomuceno F, Serna-Muñoz R, Rodríguez-
838 Llamazares S, Salido-Guadarrama I, Romero-Espinoza JA, Guadarrama-Pérez C, Sandoval-
839 Gutierrez JL, Campos F, Mondragón-Rivero EN, Ramírez-Venegas A, Castillejos-López M, Téllez-
840 Navarrete NA, Pérez-Padilla R, Vázquez-Pérez JA. Microbiota composition in the lower respiratory
841 tract is associated with severity in patients with acute respiratory distress by influenza. *Virology.* 2023
842 Feb 1;20(1):19. PMID: PMC9891757
- 843 30. Carmody LA, Kalikin LM, VanDevanter DR, Li G, Opron K, Simon RH, Caverly LJ, LiPuma JJ.
844 Changes in airway bacterial communities occur soon after initiation of antibiotic treatment of

- 845 pulmonary exacerbations in cystic fibrosis. *J Cyst Fibros Off J Eur Cyst Fibros Soc.* 2022
846 Sep;21(5):766–768. PMID: PMC10440828
- 847 31. Lloréns-Rico V, Gregory AC, Van Weyenbergh J, Jansen S, Van Buyten T, Qian J, Braz M, Menezes
848 SM, Van Mol P, Vanderbeke L, Dooms C, Gunst J, Hermans G, Meersseman P, CONTAGIOUS
849 collaborators, Wauters E, Neyts J, Lambrechts D, Wauters J, Raes J. Clinical practices underlie
850 COVID-19 patient respiratory microbiome composition and its interactions with the host. *Nat*
851 *Commun.* 2021 Oct 29;12(1):6243. PMID: PMC8556379
- 852 32. Peleg AY, Choo JM, Langan KM, Edgeworth D, Keating D, Wilson J, Rogers GB, Kotsimbos T.
853 Antibiotic exposure and interpersonal variance mask the effect of ivacaftor on respiratory microbiota
854 composition. *J Cyst Fibros Off J Eur Cyst Fibros Soc.* 2018 Jan;17(1):50–56. PMID: 29042177
- 855 33. Pittman JE, Wylie KM, Akers K, Storch GA, Hatch J, Quante J, Frayman KB, Clarke N, Davis M,
856 Stick SM, Hall GL, Montgomery G, Ranganathan S, Davis SD, Ferkol TW, Australian Respiratory
857 Early Surveillance Team for Cystic Fibrosis. Association of Antibiotics, Airway Microbiome, and
858 Inflammation in Infants with Cystic Fibrosis. *Ann Am Thorac Soc.* 2017 Oct;14(10):1548–1555.
859 PMID: PMC5718571
- 860 34. Huang YJ, Sethi S, Murphy T, Nariya S, Boushey HA, Lynch SV. Airway microbiome dynamics in
861 exacerbations of chronic obstructive pulmonary disease. *J Clin Microbiol.* United States: American
862 Society for Microbiology. All Rights Reserved; 2014;52(8):2813–2823.
- 863 35. Wang Z, Bafadhel M, Haldar K, Spivak A, Mayhew D, Miller BE, Tal-Singer R, Johnston SL,
864 Ramsheh MY, Barer MR, Brightling CE, Brown JR. Lung microbiome dynamics in COPD
865 exacerbations. *Eur Respir J.* England; 2016;47(4):1082–1092.
- 866 36. Zakharkina T, Martin-Loeches I, Matamoros S, Pova A, Torres A, Kastelijn JB, Hofstra JJ, de
867 Wever B, de Jong M, Schultz MJ, Sterk PJ, Artigas A, Bos LDJ. The dynamics of the pulmonary

- 868 microbiome during mechanical ventilation in the intensive care unit and the association with
869 occurrence of pneumonia. *Thorax*. 2017 Sep;72(9):803–810. PMID: 28100714
- 870 37. Panzer AR, Lynch SV, Langelier C, Christie JD, McCauley K, Nelson M, Cheung CK, Benowitz
871 NL, Cohen MJ, Calfee CS. Lung Microbiota Is Related to Smoking Status and to Development of
872 Acute Respiratory Distress Syndrome in Critically Ill Trauma Patients. *Am J Respir Crit Care Med*.
873 2018 Mar 1;197(5):621–631. PMID: 29605235
- 874 38. Madaras-Kelly K, Jones M, Remington R, Hill N, Huttner B, Samore M. Development of an
875 antibiotic spectrum score based on veterans affairs culture and susceptibility data for the purpose of
876 measuring antibiotic de-escalation: a modified Delphi approach. *Infect Control Hosp Epidemiol*.
877 2014 Sep;35(9):1103–1113. PMID: 25078427
- 878 39. Chanderraj R, Baker JM, Kay SG, Brown CA, Hinkle KJ, Fergle DJ, McDonald RA, Falkowski NR,
879 Metcalf JD, Kaye KS, Woods RJ, Prescott HC, Sjoding MW, Dickson RP. In critically ill patients,
880 anti-anaerobic antibiotics increase risk of adverse clinical outcomes. *Eur Respir J*. 2023
881 Feb;61(2):2200910. PMID: 36909213
- 882 40. Ogimi C, Krantz EM, Golob JL, Waghmare A, Liu C, Leisenring WM, Woodard CR, Marquis S,
883 Kuypers JM, Jerome KR, Pergam SA, Fredricks DN, Sorrow ML, Englund JA, Boeckh M. Antibiotic
884 Exposure Prior to Respiratory Viral Infection Is Associated with Progression to Lower Respiratory
885 Tract Disease in Allogeneic Hematopoietic Cell Transplant Recipients. *Biol Blood Marrow
886 Transplant J Am Soc Blood Marrow Transplant*. United States: The American Society for Blood and
887 Marrow Transplantation. Published by Elsevier Inc; 2018;24(11):2293–2301.
- 888 41. Tanaka JS, Young RR, Heston SM, Jenkins K, Spees LP, Sung AD, Corbet K, Thompson JC,
889 Bohannon L, Martin PL, Stokhuyzen A, Vinesett R, Ward DV, Bhattarai SK, Bucci V, Arshad M,
890 Seed PC, Kelly MS. Anaerobic Antibiotics and the Risk of Graft-versus-Host Disease after

- 891 Allogeneic Hematopoietic Stem Cell Transplantation. *Biol Blood Marrow Transplant J Am Soc*
892 *Blood Marrow Transplant*. 2020 Nov;26(11):2053–2060. PMID: PMC7609492
- 893 42. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani
894 MS, Luca BA, Steiner D, Diehn M, Alizadeh AA. Determining cell type abundance and expression
895 from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019 Jul;37(7):773–782. PMID:
896 PMC6610714
- 897 43. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling Cell Type Abundance and Expression in
898 Bulk Tissues with CIBERSORTx. *Methods Mol Biol Clifton NJ*. 2020;2117:135–157. PMID:
899 PMC7695353
- 900 44. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley SD, Mori Y,
901 Seita J, Berry GJ, Shrager JB, Metzger RJ, Kuo CS, Neff N, Weissman IL, Quake SR, Krasnow
902 MA. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*. 2020
903 Nov;587(7835):619–625. PMID: PMC7704697
- 904 45. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, D’Agostino R, Castro M, Curran-
905 Everett D, Fitzpatrick AM, Gaston B, Jarjour NN, Sorkness R, Calhoun WJ, Chung KF, Comhair
906 SAA, Dweik RA, Israel E, Peters SP, Busse WW, Erzurum SC, Bleeker ER, National Heart, Lung,
907 and Blood Institute’s Severe Asthma Research Program. Identification of asthma phenotypes using
908 cluster analysis in the Severe Asthma Research Program. *Am J Respir Crit Care Med*. 2010 Feb
909 15;181(4):315–323. PMID: PMC2822971
- 910 46. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, NHLBI ARDS
911 Network. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from
912 two randomised controlled trials. *LancetRespiratory Med*. England: Elsevier Ltd; 2014;2(8):611–
913 620.

- 914 47. Calfee CS, Delucchi KL, Sinha P, Matthay MA, Hackett J, Shankar-Hari M, McDowell C, Laffey
915 JG, O’Kane CM, McAuley DF, Irish Critical Care Trials Group. Acute respiratory distress syndrome
916 subphenotypes and differential response to simvastatin: secondary analysis of a randomised
917 controlled trial. *Lancet Respir Med*. 2018 Sep;6(9):691–698. PMID: PMC6201750
- 918 48. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, Curtis
919 JL. Bacterial Topography of the Healthy Human Lower Respiratory Tract. *mBio*. United States:
920 Dickson et al; 2017;8(1):10.1128/mBio.02287-16.
- 921 49. Di Simone SK, Rudloff I, Nold-Petry CA, Forster SC, Nold MF. Understanding respiratory
922 microbiome-immune system interactions in health and disease. *Sci Transl Med*. 2023 Jan
923 11;15(678):eabq5126. PMID: 36630485
- 924 50. Yao Y, Jeyanathan M, Haddadi S, Barra NG, Vaseghi-Shanjani M, Damjanovic D, Lai R, Afkhami
925 S, Chen Y, Dvorkin-Gheva A, Robbins CS, Schertzer JD, Xing Z. Induction of Autonomous
926 Memory Alveolar Macrophages Requires T Cell Help and Is Critical to Trained Immunity. *Cell*.
927 2018 Nov 29;175(6):1634-1650.e17. PMID: 30433869
- 928 51. Niec RE, Rudensky AY, Fuchs E. Inflammatory adaptation in barrier tissues. *Cell*. 2021 Jun
929 24;184(13):3361–3375. PMID: PMC8336675
- 930 52. Schupp JC, Khanal S, Gomez JL, Sauler M, Adams TS, Chupp GL, Yan X, Poli S, Zhao Y,
931 Montgomery RR, Rosas IO, Dela Cruz CS, Bruscia EM, Egan ME, Kaminski N, Britto CJ. Single-
932 Cell Transcriptional Archetypes of Airway Inflammation in Cystic Fibrosis. *Am J Respir Crit Care*
933 *Med*. 2020 Nov 15;202(10):1419–1429. PMID: PMC7667912
- 934 53. Margaroli C, Garratt LW, Horati H, Dittrich AS, Rosenow T, Montgomery ST, Frey DL, Brown
935 MR, Schultz C, Guglani L, Kicic A, Peng L, Scholte BJ, Mall MA, Janssens HM, Stick SM,
936 Tirouvanziam R. Elastase Exocytosis by Airway Neutrophils Is Associated with Early Lung Damage

- 937 in Children with Cystic Fibrosis. *Am J Respir Crit Care Med*. 2019 Apr 1;199(7):873–881. PMID:
938 PMC6444666
- 939 54. Segal LN, Clemente JC, Tsay JCJ, Koralov SB, Keller BC, Wu BG, Li Y, Shen N, Ghedin E, Morris
940 A, Diaz P, Huang L, Wikoff WR, Ubeda C, Artacho A, Rom WN, Serman DH, Collman RG, Blaser
941 MJ, Weiden MD. Enrichment of the lung microbiome with oral taxa is associated with lung
942 inflammation of a Th17 phenotype. *Nat Microbiol*. 2016 Apr 4;1:16031. PMID: PMC5010013
- 943 55. Sulaiman I, Wu BG, Chung M, Isaacs B, Tsay JCJ, Holub M, Barnett CR, Kwok B, Kugler MC,
944 Natalini JG, Singh S, Li Y, Schluger R, Carpenito J, Collazo D, Perez L, Kyeremateng Y, Chang M,
945 Campbell CD, Hansbro PM, Oppenheimer BW, Berger KI, Goldring RM, Koralov SB, Weiden MD,
946 Xiao R, D’Armiento J, Clemente JC, Ghedin E, Segal LN. Lower Airway Dysbiosis Augments Lung
947 Inflammatory Injury in Mild-to-Moderate COPD. *Am J Respir Crit Care Med*. 2023 Sep 7; PMID:
948 37677136
- 949 56. Burgos da Silva M, Ponce DM, Dai A, M Devlin S, Gomes ALC, Moore G, Slingerland J, Shouval
950 R, Armijo GK, DeWolf S, Fei T, Clurman A, Fontana E, Amoretti LA, Wright RJ, Androva H,
951 Miltiadous O, Perales MA, Taur Y, Peled JU, van den Brink MRM. Preservation of the fecal
952 microbiome is associated with reduced severity of graft-versus-host disease. *Blood*. 2022 Dec
953 1;140(22):2385–2397. PMID: PMC9837450
- 954 57. Peled JU, Gomes ALC, Devlin SM, Littmann ER, Taur Y, Sung AD, Weber D, Hashimoto D,
955 Slingerland AE, Slingerland JB, Maloy M, Clurman AG, Stein-Thoeringer CK, Markey KA,
956 Docampo MD, Burgos da Silva M, Khan N, Gessner A, Messina JA, Romero K, Lew MV, Bush A,
957 Bohannon L, Brereton DG, Fontana E, Amoretti LA, Wright RJ, Armijo GK, Shono Y, Sanchez-
958 Escamilla M, Castillo Flores N, Alarcon Tomas A, Lin RJ, Yanez San Segundo L, Shah GL, Cho C,
959 Scordo M, Politikos I, Hayasaka K, Hasegawa Y, Gyurkocza B, Ponce DM, Barker JN, Perales MA,
960 Giralt SA, Jenq RR, Teshima T, Chao NJ, Holler E, Xavier JB, Pamer EG, van den Brink MRM.

- 961 Microbiota as Predictor of Mortality in Allogeneic Hematopoietic-Cell Transplantation. *N Engl J*
962 *Med.* United States: Massachusetts Medical Society; 2020;382(9):822–834.
- 963 58. Shono Y, Docampo MD, Peled JU, Perobelli SM, Velardi E, Tsai JJ, Slingerland AE, Smith OM,
964 Young LF, Gupta J, Lieberman SR, Jay HV, Ahr KF, Porosnicu Rodriguez KA, Xu K, Calarfiore
965 M, Poeck H, Caballero S, Devlin SM, Rapaport F, Dudakov JA, Hanash AM, Gyurkocza B, Murphy
966 GF, Gomes C, Liu C, Moss EL, Falconer SB, Bhatt AS, Taur Y, Pamer EG, van den Brink MRM,
967 Jenq RR. Increased GVHD-related mortality with broad-spectrum antibiotic use after allogeneic
968 hematopoietic stem cell transplantation in human patients and mice. *Sci Transl Med.* United States:
969 American Association for the Advancement of Science; 2016;8(339):339ra71.
- 970 59. O’Dwyer DN, Zhou X, Wilke CA, Xia M, Falkowski NR, Norman KC, Arnold KB, Huffnagle GB,
971 Murray S, Erb-Downward JR, Yanik GA, Moore BB, Dickson RP. Lung Dysbiosis, Inflammation,
972 and Injury in Hematopoietic Cell Transplantation. *Am J Respir Crit Care Med.* United States;
973 2018;198(10):1312–1321.
- 974 60. Abreu NA, Nagalingam NA, Song Y, Roediger FC, Pletcher SD, Goldberg AN, Lynch SV. Sinus
975 microbiome diversity depletion and *Corynebacterium tuberculostearicum* enrichment mediates
976 rhinosinusitis. *Sci Transl Med.* United States; 2012;4(151):151ra124.
- 977 61. Goeteyn E, Grassi L, Van den Bossche S, Rigauts C, Vande Weygaerde Y, Van Braeckel E, Maes
978 T, Bracke KR, Crabbé A. Commensal bacteria of the lung microbiota synergistically inhibit
979 inflammation in a three-dimensional epithelial cell model. *Front Immunol.* 2023;14:1176044.
980 PMID: PMC10164748
- 981 62. Rigauts C, Aizawa J, Taylor SL, Rogers GB, Govaerts M, Cos P, Ostyn L, Sims S, Vandeplassche
982 E, Sze M, Dondelinger Y, Vereecke L, Van Acker H, Simpson JL, Burr L, Willems A, Tunney MM,
983 Cigana C, Bragonzi A, Coenye T, Crabbé A. *Rothia mucilaginosa* is an anti-inflammatory bacterium

- 984 in the respiratory tract of patients with chronic lung disease. *Eur Respir J*. 2022 May;59(5):2101293.
985 PMID: PMC9068977
- 986 63. Brown RL, Sequeira RP, Clarke TB. The microbiota protects against respiratory infection via GM-
987 CSF signaling. *Nat Commun*. 2017 Nov 15;8(1):1512. PMID: PMC5688119
- 988 64. Horn KJ, Schopper MA, Drigot ZG, Clark SE. Airway Prevotella promote TLR2-dependent
989 neutrophil activation and rapid clearance of *Streptococcus pneumoniae* from the lung. *Nat Commun*.
990 2022 Jun 9;13(1):3321. PMID: PMC9184549
- 991 65. Wu BG, Sulaiman I, Tsay JCJ, Perez L, Franca B, Li Y, Wang J, Gonzalez AN, El-Ashmawy M,
992 Carpenito J, Olsen E, Sauthoff M, Yie K, Liu X, Shen N, Clemente JC, Kapoor B, Zangari T,
993 Mezzano V, Loomis C, Weiden MD, Koralov SB, D'Armiento J, Ahuja SK, Wu XR, Weiser JN,
994 Segal LN. Episodic Aspiration with Oral Commensals Induces a MyD88-dependent, Pulmonary T-
995 Helper Cell Type 17 Response that Mitigates Susceptibility to *Streptococcus pneumoniae*. *Am J*
996 *Respir Crit Care Med*. 2021 May 1;203(9):1099–1111. PMID: PMC8314894
- 997 66. Yan Z, Chen B, Yang Y, Yi X, Wei M, Ecklu-Mensah G, Buschmann MM, Liu H, Gao J, Liang W,
998 Liu X, Yang J, Ma W, Liang Z, Wang F, Chen D, Wang L, Shi W, Stampfli MR, Li P, Gong S,
999 Chen X, Shu W, El-Omar EM, Gilbert JA, Blaser MJ, Zhou H, Chen R, Wang Z. Multi-omics
1000 analyses of airway host-microbe interactions in chronic obstructive pulmonary disease identify
1001 potential therapeutic interventions. *Nat Microbiol*. 2022 Sep;7(9):1361–1375. PMID: 35995842
- 1002 67. Liang W, Yang Y, Gong S, Wei M, Ma Y, Feng R, Gao J, Liu X, Tu F, Ma W, Yi X, Liang Z, Wang
1003 F, Wang L, Chen D, Shu W, Miller BE, Tal-Singer R, Donaldson GC, Wedzicha JA, Singh D,
1004 Wilkinson TMA, Brightling CE, Chen R, Zhong N, Wang Z. Airway dysbiosis accelerates lung
1005 function decline in chronic obstructive pulmonary disease. *Cell Host Microbe*. 2023 Jun
1006 14;31(6):1054-1070.e9. PMID: 37207649

- 1007 68. Peleg AY, Hogan DA, Mylonakis E. Medically important bacterial-fungal interactions. *Nat Rev*
1008 *Microbiol.* 2010 May;8(5):340–349. PMID: 20348933
- 1009 69. Rao C, Coyte KZ, Bainter W, Geha RS, Martin CR, Rakoff-Nahoum S. Multi-kingdom ecological
1010 drivers of microbiota assembly in preterm infants. *Nature.* 2021 Mar;591(7851):633–638. PMID:
1011 PMC7990694
- 1012 70. van Tilburg Bernardes E, Pettersen VK, Gutierrez MW, Laforest-Lapointe I, Jendzjowsky NG,
1013 Cavin JB, Vicentini FA, Keenan CM, Ramay HR, Samara J, MacNaughton WK, Wilson RJA, Kelly
1014 MM, McCoy KD, Sharkey KA, Arrieta MC. Intestinal fungi are causally implicated in microbiome
1015 assembly and immune development in mice. *Nat Commun.* 2020 May 22;11(1):2577. PMID:
1016 PMC7244730
- 1017 71. Rolling T, Zhai B, Gjonbalaj M, Tosini N, Yasuma-Mitobe K, Fontana E, Amoretti LA, Wright RJ,
1018 Ponce DM, Perales MA, Xavier JB, van den Brink MRM, Markey KA, Peled JU, Taur Y, Hohl TM.
1019 Haematopoietic cell transplantation outcomes are linked to intestinal mycobiota dynamics and an
1020 expansion of *Candida parapsilosis* complex species. *Nat Microbiol.* 2021 Dec;6(12):1505–1515.
1021 PMID: PMC8939874
- 1022 72. Yang YT, Wong D, Ashcroft DM, Massey J, MacKenna B, Fisher L, Mehrkar A, Bacon SC,
1023 OpenSAFELY collaborative, Hand K, Zhong X, Fahmi A, Goldacre B, van Staa T, Palin V.
1024 Repeated antibiotic exposure and risk of hospitalisation and death following COVID-19 infection
1025 (OpenSAFELY): a matched case-control study. *EClinicalMedicine.* 2023 Jul;61:102064. PMID:
1026 PMC10388579
- 1027 73. Shouval R, Waters NR, Gomes ALC, Zuanelli Brambilla C, Fei T, Devlin SM, Nguyen CL, Markey
1028 KA, Dai A, Slingerland JB, Clurman AG, Fontana E, Amoretti LA, Wright RJ, Hohl TM, Taur Y,
1029 Sung AD, Weber D, Hashimoto D, Teshima T, Chao NJ, Holler E, Scordo M, Giralt SA, Perales

- 1030 MA, Peled JU, van den Brink MRM. Conditioning Regimens are Associated with Distinct Patterns
1031 of Microbiota Injury in Allogeneic Hematopoietic Cell Transplantation. *Clin Cancer Res Off J Am*
1032 *Assoc Cancer Res.* 2023 Jan 4;29(1):165–173. PMID: PMC9812902
- 1033 74. Charalampous T, Aloclea-Medina A, Snell LB, Alder C, Tan M, Williams TGS, Al-Yaakoubi N,
1034 Humayun G, Meadows CIS, Wyncoll DLA, Paul R, Hemsley CJ, Jeyaratnam D, Newsholme W,
1035 Goldenberg S, Patel A, Tucker F, Nebbia G, Wilks M, Chand M, Cliff PR, Batra R, O’Grady J,
1036 Barrett NA, Edgeworth JD. Routine Metagenomics Service for Intensive Care Unit Patients with
1037 Respiratory Infection. *Am J Respir Crit Care Med.* 2023 Nov 8; PMID: 37938162
- 1038 75. Chotirmall SH, Bogaert D, Chalmers JD, Cox MJ, Hansbro PM, Huang YJ, Molyneaux PL,
1039 O’Dwyer DN, Pragman AA, Rogers GB, Segal LN, Dickson RP. Therapeutic Targeting of the
1040 Respiratory Microbiome. *Am J Respir Crit Care Med.* 2022 Sep 1;206(5):535–544. PMID:
1041 PMC9716896
- 1042 76. Wilson MR, Shanbhag NM, Reid MJ, Singhal NS, Gelfand JM, Sample HA, Benkli B, O’Donovan
1043 BD, Ali IK, Keating MK, Dunnebacke TH, Wood MD, Bollen A, DeRisi JL. Diagnosing *Balamuthia*
1044 *mandrillaris* Encephalitis With Metagenomic Deep Sequencing. *Ann Neurol.* United States: The
1045 Authors Annals of Neurology published by Wiley Periodicals, Inc. on behalf of American
1046 Neurological Association; 2015;78(5):722–730.
- 1047 77. Doan T, Wilson MR, Crawford ED, Chow ED, Khan LM, Knopp KA, O’Donovan BD, Xia D,
1048 Hacker JK, Stewart JM, Gonzales JA, Acharya NR, DeRisi JL. Illuminating uveitis: metagenomic
1049 deep sequencing identifies common and rare pathogens. *Genome Med.* England; 2016;8(1):90–6.
- 1050 78. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S,
1051 Federman S, Miller S, Sokolic R, Garabedian E, Candotti F, Buckley RH, Reed KD, Meyer TL,
1052 Seroogy CM, Galloway R, Henderson SL, Gern JE, DeRisi JL, Chiu CY. Actionable diagnosis of

- 1053 neuroleptospirosis by next-generation sequencing. *N Engl J Med*. United States;
1054 2014;370(25):2408–2417.
- 1055 79. Gurczynski SJ, Zhou X, Flaherty M, Wilke CA, Moore BB. Bone marrow transplant-induced
1056 alterations in Notch signaling promote pathologic Th17 responses to γ -herpesvirus infection.
1057 *Mucosal Immunol*. 2018 May;11(3):881–893. PMID: PMC5906203
- 1058 80. Zinter MS, Hume JR. Effects of Hematopoietic Cell Transplantation on the Pulmonary Immune
1059 Response to Infection. *Front Pediatr*. 2021;9:634566. PMID: PMC7871005
- 1060 81. Zhou X, Moore BB. Experimental Models of Infectious Pulmonary Complications Following
1061 Hematopoietic Cell Transplantation. *Front Immunol*. 2021;12:718603. PMID: PMC8415416
- 1062 82. Domingo-Gonzalez R, Martínez-Colón GJ, Smith AJ, Smith CK, Ballinger MN, Xia M, Murray S,
1063 Kaplan MJ, Yanik GA, Moore BB. Inhibition of Neutrophil Extracellular Trap Formation after Stem
1064 Cell Transplant by Prostaglandin E2. *Am J Respir Crit Care Med*. 2016 Jan 15;193(2):186–197.
1065 PMID: PMC4731709
- 1066 83. Matthaïou EI, Sharifi H, O'Donnell C, Chiu W, Owyang C, Chatterjee P, Turk I, Johnston L,
1067 Brondstetter T, Morris K, Cheng GS, Hsu JL. The safety and tolerability of pirfenidone for
1068 bronchiolitis obliterans syndrome after hematopoietic cell transplant (STOP-BOS) trial. *Bone
1069 Marrow Transplant*. 2022 Aug;57(8):1319–1326. PMID: PMC9357121
- 1070 84. Liu Q, Tian X, Maruyama D, Arjomandi M, Prakash A. Lung immune tone via gut-lung axis: gut-
1071 derived LPS and short-chain fatty acids' immunometabolic regulation of lung IL-1 β , FFAR2, and
1072 FFAR3 expression. *Am J Physiol Lung Cell Mol Physiol*. 2021 Jul 1;321(1):L65–L78. PMID:
1073 PMC8321849

- 1074 85. Josefsdottir KS, Baldrige MT, Kadmon CS, King KY. Antibiotics impair murine hematopoiesis by
1075 depleting the intestinal microbiota. *Blood*. 2017 Feb 9;129(6):729–739. PMID: PMC5301822
- 1076 86. Faro A, Wood RE, Schechter MS, Leong AB, Wittkugel E, Abode K, Chmiel JF, Daines C, Davis
1077 S, Eber E, Huddleston C, Kilbaugh T, Kurland G, Midulla F, Molter D, Montgomery GS, Retsch-
1078 Bogart G, Rutter MJ, Visner G, Walczak SA, Ferkol TW, Michelson PH, American Thoracic Society
1079 Ad Hoc Committee on Flexible Airway Endoscopy in Children. Official American Thoracic Society
1080 technical standards: flexible airway endoscopy in children. *Am J Respir Crit Care Med*. 2015 May
1081 1;191(9):1066–1080. PMID: 25932763
- 1082 87. Radhakrishnan D, Yamashita C, Gillio-Meina C, Fraser DD. Translational research in pediatrics III:
1083 bronchoalveolar lavage. *Pediatrics*. United States: by the American Academy of Pediatrics;
1084 2014;134(1):135–154.
- 1085 88. Martinu T, Koutsokera A, Benden C, Cantu E, Chambers D, Cypel M, Edelman J, Emtiazjoo A,
1086 Fisher AJ, Greenland JR, Hayes D, Hwang D, Keller BC, Lease ED, Perch M, Sato M, Todd JL,
1087 Verleden S, von der Thüsen J, Weigt SS, Keshavjee S, bronchoalveolar lavage standardization
1088 workgroup. International Society for Heart and Lung Transplantation consensus statement for the
1089 standardization of bronchoalveolar lavage in lung transplantation. *J Heart Lung Transplant Off Publ
1090 Int Soc Heart Transplant*. 2020 Nov;39(11):1171–1190. PMID: PMC7361106
- 1091 89. Kim J, Kim MS, Koh AY, Xie Y, Zhan X. FMAP: Functional Mapping and Analysis Pipeline for
1092 metagenomics and metatranscriptomics studies. *BMC Bioinformatics*. 2016 Oct 10;17(1):420.
1093 PMID: PMC5057277
- 1094 90. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat
1095 Methods*. 2015 Jan;12(1):59–60. PMID: 25402007

- 1096 91. Abubucker S, Segata N, Goll J, Schubert AM, IZard J, Cantarel BL, Rodriguez-Mueller B, Zucker
1097 J, Thiagarajan M, Henrissat B, White O, Kelley ST, Methé B, Schloss PD, Gevers D, Mitreva M,
1098 Huttenhower C. Metabolic reconstruction for metagenomic data and its application to the human
1099 microbiome. *PLoS Comput Biol.* 2012;8(6):e1002358. PMID: PMC3374609
- 1100 92. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. UniRef clusters: a
1101 comprehensive and scalable alternative for improving sequence similarity searches. *Bioinforma Oxf*
1102 *Engl.* 2015 Mar 15;31(6):926–932. PMID: PMC4375400
- 1103 93. Tanabe M, Kanehisa M. Using the KEGG database resource. *Curr Protoc Bioinforma.* 2012
1104 Jun;Chapter 1:1.12.1-1.12.43. PMID: 22700311
- 1105 94. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of
1106 large-scale molecular data sets. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D109-114.
1107 PMID: PMC3245020
- 1108 95. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W, Stegle O.
1109 Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets.
1110 *Mol Syst Biol.* 2018 Jun 20;14(6):e8124. PMID: PMC6010767
- 1111 96. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq
1112 data with DESeq2. *Genome Biol.* 2014;15(12):550. PMID: PMC4302049
- 1113 97. Oksanen J, Weedon J. *vegan: Community Ecology Package* [Internet]. 2002. Available from:
1114 <https://CRAN.R-project.org/package=vegan>
- 1115 98. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.*
1116 *England;* 2010;11(10):R106.

- 1117 99. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation
1118 Analysis. *J Stat Softw.* 2014 Sep 2;59:1–38.
- 1119 100. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. *Psychol Methods.*
1120 United States: APA, all rights reserved; 2010;15(4):309–334.
- 1121 101. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-
1122 seq data. *BMC Bioinformatics.* 2013 Jan 16;14:7. PMID: PMC3618321
- 1123 102. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, Griss J, Sevilla C,
1124 Matthews L, Gong C, Deng C, Varusai T, Ragueneau E, Haider Y, May B, Shamovsky V, Weiser
1125 J, Brunson T, Sanati N, Beckman L, Shao X, Fabregat A, Sidiropoulos K, Murillo J, Viteri G, Cook
1126 J, Shorser S, Bader G, Demir E, Sander C, Haw R, Wu G, Stein L, Hermjakob H, D’Eustachio P.
1127 The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 2022 Jan 7;50(D1):D687–D692.
1128 PMID: PMC8689983
- 1129 103. Mandric I, Rotman J, Yang HT, Strauli N, Montoya DJ, Van Der Wey W, Ronas JR, Statz B, Yao
1130 D, Petrova V, Zelikovsky A, Spreafico R, Shifman S, Zaitlen N, Rossetti M, Ansel KM, Eskin E,
1131 Mangul S. Profiling immunoglobulin repertoires across multiple human tissues using RNA
1132 sequencing. *Nat Commun.* 2020 Jun 19;11(1):3126. PMID: PMC7305308
- 1133 104. Peng K, Nowicki TS, Campbell K, Vahed M, Peng D, Meng Y, Nagareddy A, Huang YN, Karlsberg
1134 A, Miller Z, Brito J, Nadel B, Pak VM, Abedalthagafi MS, Burkhardt AM, Alachkar H, Ribas A,
1135 Mangul S. Rigorous benchmarking of T-cell receptor repertoire profiling methods for cancer RNA
1136 sequencing. *Brief Bioinform.* 2023 Jul 20;24(4):bbad220. PMID: PMC10359085
- 1137 105. Breiman L. Random Forests. *Mach Learn.* 2001;45:5–32.

- 1138 106. Athey S, Tibshirani J, Wager S. Generalized random forests. *Ann Stat. Institute of Mathematical*
1139 *Statistics*; 2019 Apr;47(2):1148–1178.

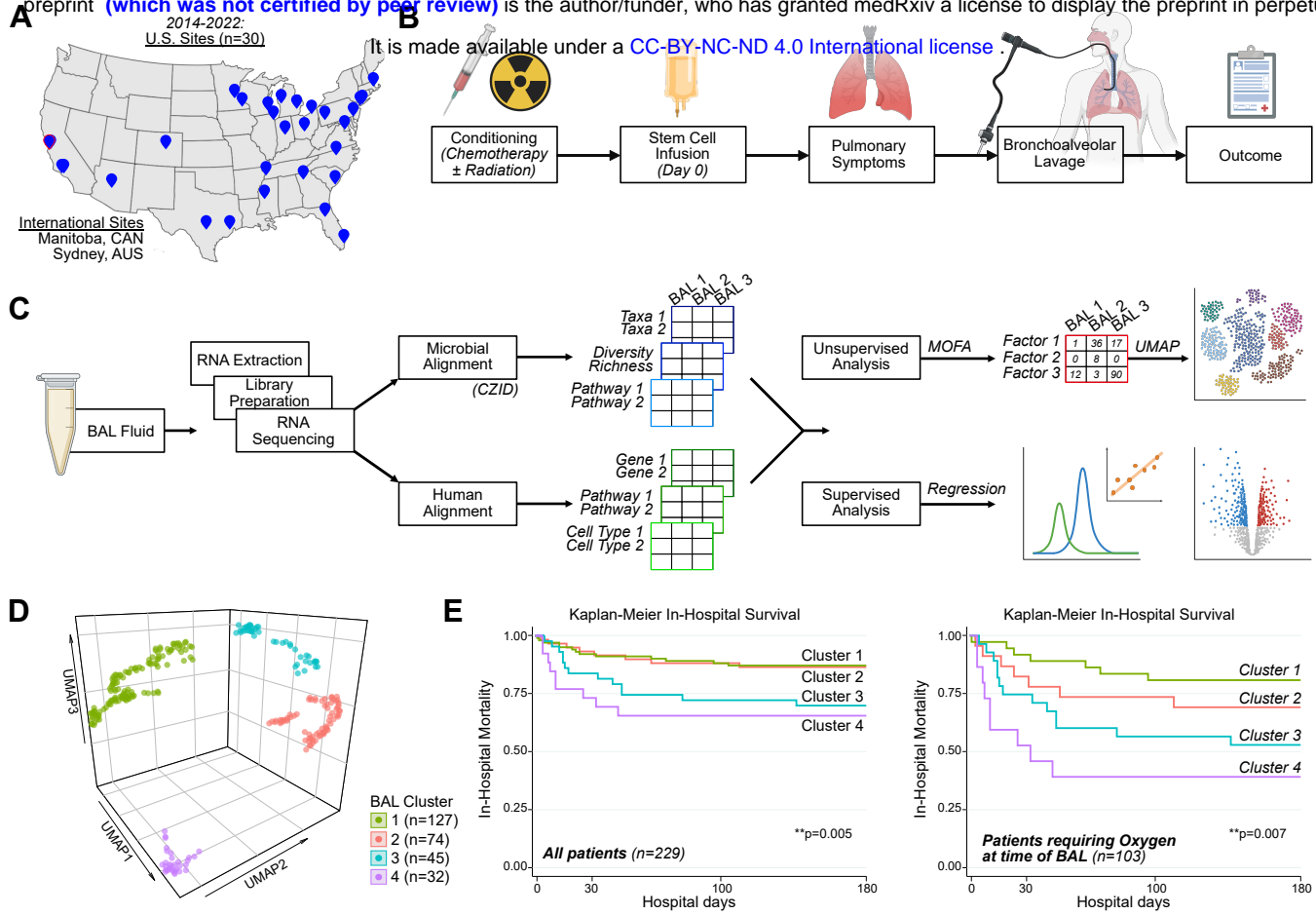


Figure 1. Study design and clinical outcomes. (A) Patients were recruited from 32 participating children’s hospitals in the United States, Canada, and Australia. (B) Study design concept diagram. (C) BAL processing and analysis workflow. (D) Four microbiome-transcriptome clusters were identified. (E) In-hospital survival for all patients (left) and the subset requiring respiratory support prior to testing (right) was plotted according to BAL cluster and differences were analyzed with the log rank test.

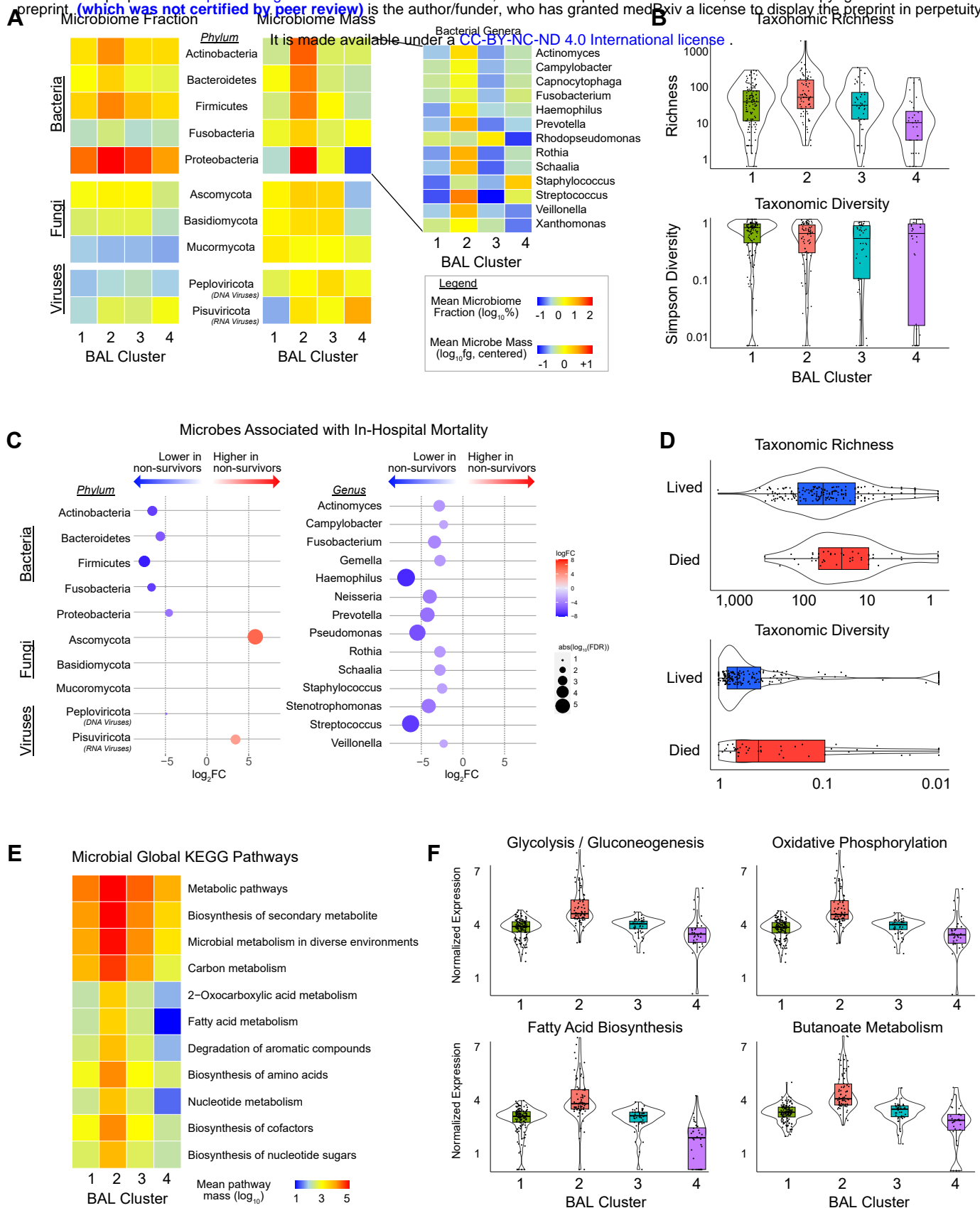
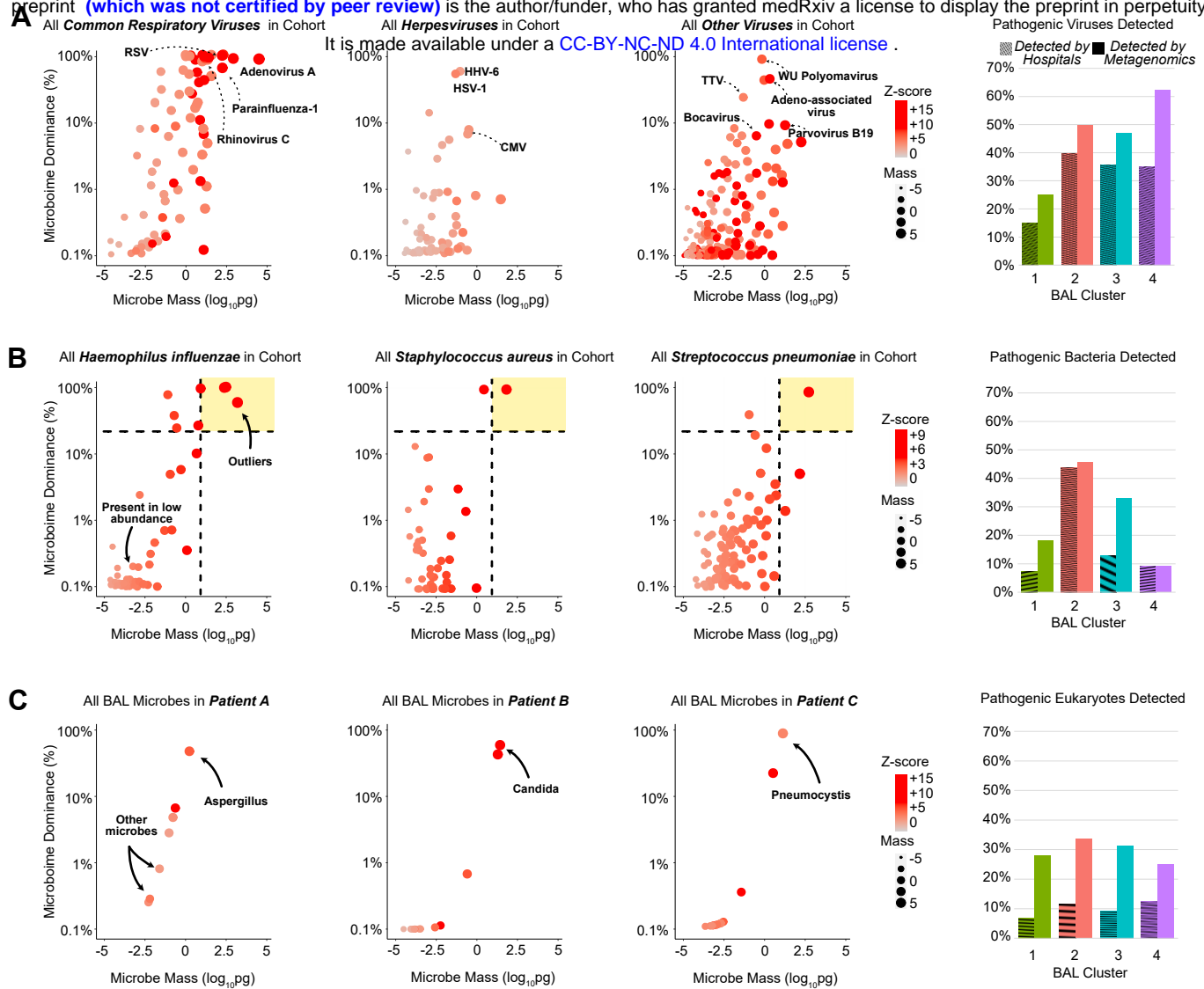


Figure 2. BAL microbiome. (A) The fraction (left) and mass (right) of major bacterial, viral, and fungal phyla are plotted, with shading representing the average for each of the 4 BAL clusters. The average mass of bacterial genera in each of the 4 BAL clusters are shown to the right. (B) Taxonomic richness and diversity are plotted across the 4 BAL clusters. (C) Microbes associated with in-hospital mortality were identified using negative binomial generalized linear models (*edgeR*) and are plotted according to \log_2FC (position, color) and FDR (dot size). (D) Taxonomic richness and diversity stratified by survival status. (E) Microbial alignments to KEGG metabolic pathways were averaged for each BAL cluster. (F) Select metabolic pathways that differ across the BAL clusters are shown.



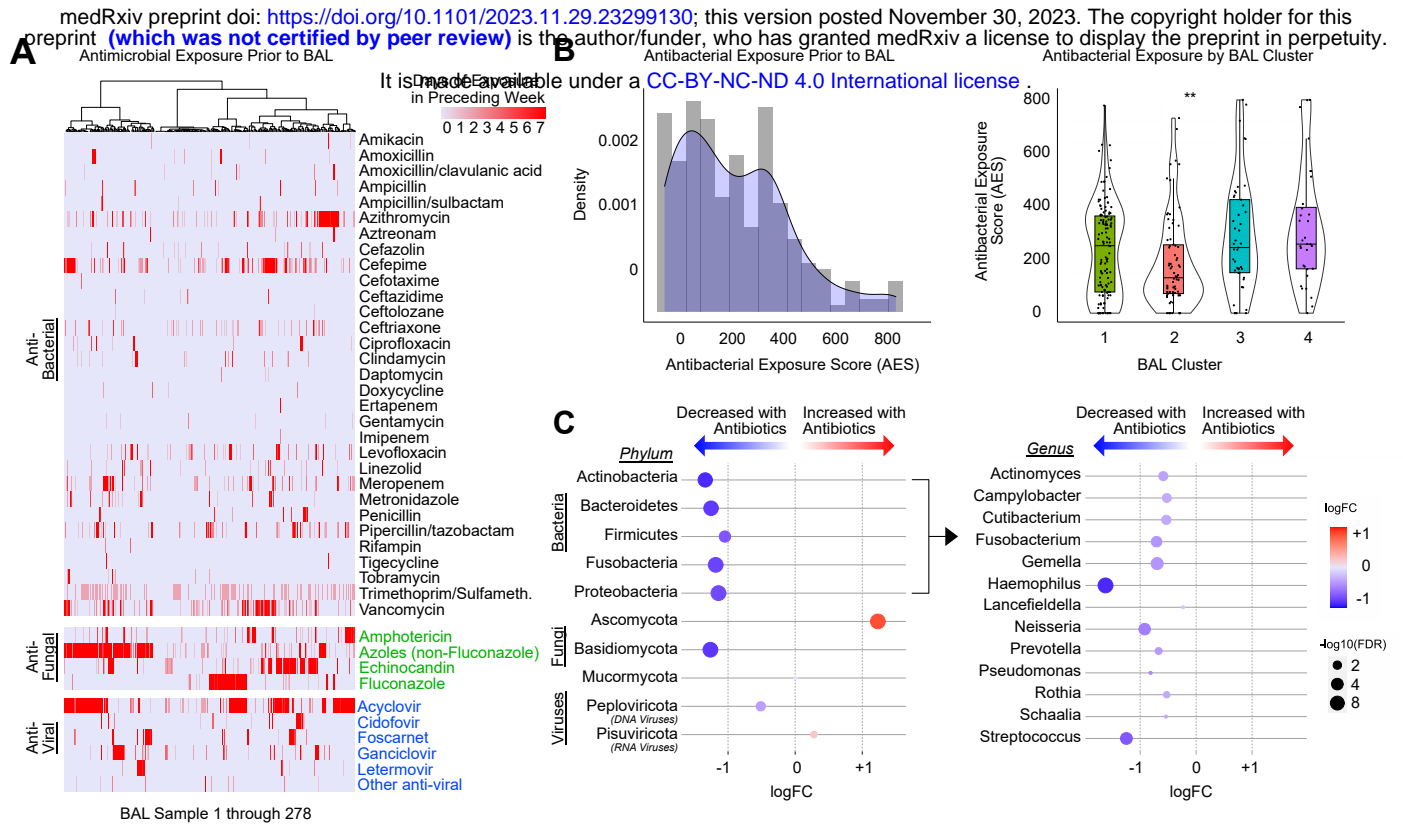


Figure 4. Antibiotic exposure and impact on BAL microbiome. (A) Days of antimicrobials are listed for antibacterials (black), antifungals (green), and antivirals (blue). Patients are listed in columns and shading indicates number of days of exposure to each antibiotic in the week preceding BAL. (B) Antibiotic exposure score (AES) was calculated prior to each BAL as the sum of antibiotic exposure days times a broadness weighting factor, summed for all therapies received in the week preceding BAL. AES varied across the clusters and was highest for patients in Cluster 4. (C) Negative binomial generalized linear models were used to test for BAL microbes associated with AES. Microbes are listed in rows, with phyla shown on the left and bacterial genera shown on the right.

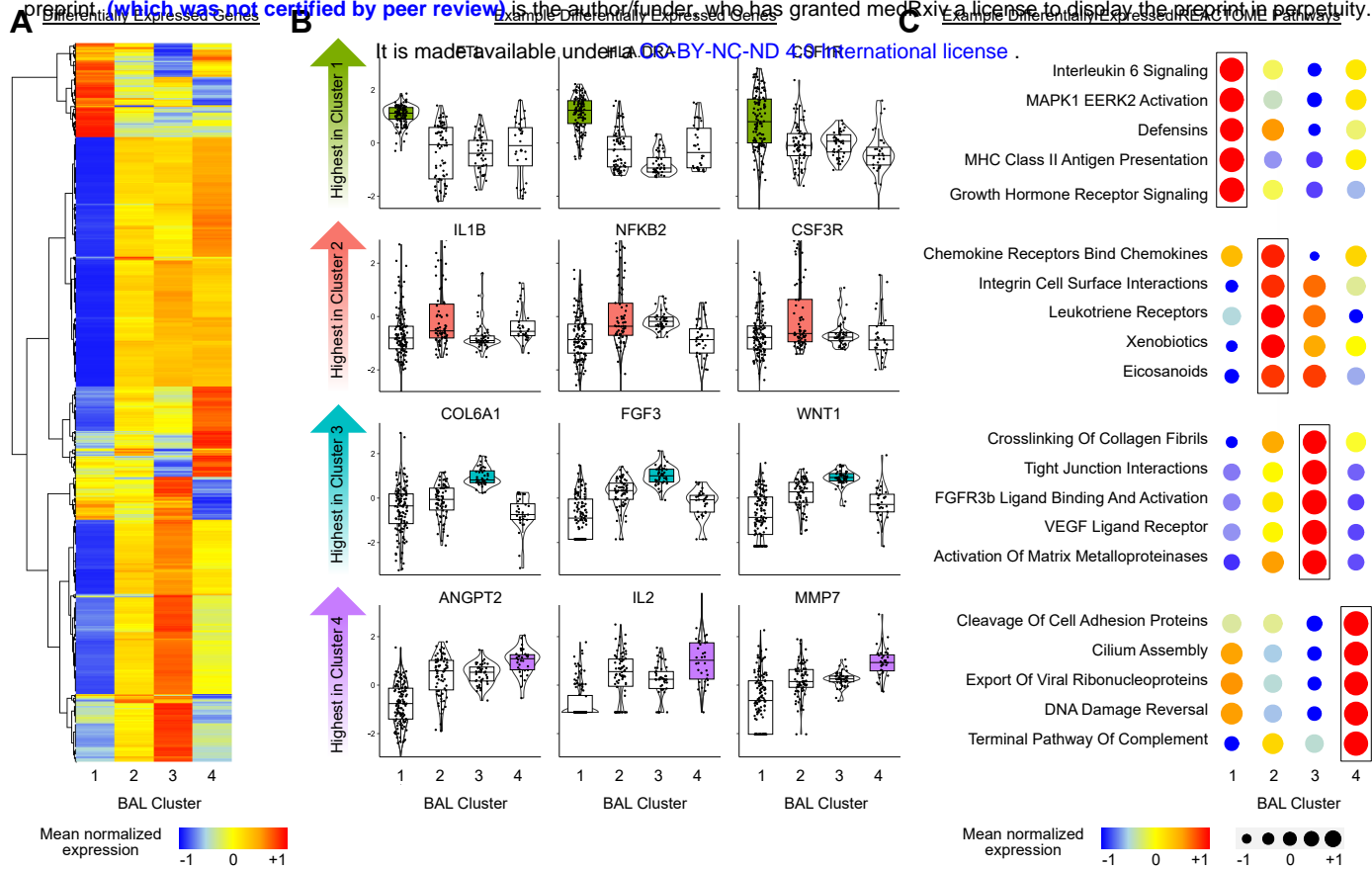


Figure 5. BAL gene expression. (A) Differentially expressed genes were identified by 4-way ANOVA like analysis with negative binomial generalized linear models. Mean normalized expression levels for significant genes are displayed for the 4 BAL clusters. (B) Individual differentially expressed genes were identified across the 4 clusters (edgeR) and variance-stabilized transformed gene counts for select genes highest in each of the 4 clusters are plotted. (C) Gene set enrichment scores to Reactome pathways were calculated and example gene sets most enriched in each of the 4 clusters are shown.