

TRANSPLANTATION

Pulmonary microbiome and gene expression signatures differentiate lung function in pediatric hematopoietic cell transplant candidates

Matt S. Zinter^{1,2*†}, A. Birgitta Versluys^{3,4†}, Caroline A. Lindemans^{3,4}, Madeline Y. Mayday⁵, Gustavo Reyes¹, Sara Sunshine⁶, Marilynn Chan⁷, Elizabeth K. Fiorino⁸, Maria Cancio^{9,10}, Sabine Prevaes¹¹, Marina Sirota^{12,13}, Michael A. Matthay¹⁴, Sandhya Kharbanda², Christopher C. Dvorak², Jaap J. Boelens^{9,10‡}, Joseph L. DeRisi^{6,15‡}

Impaired baseline lung function is associated with mortality after pediatric allogeneic hematopoietic cell transplantation (HCT), yet limited knowledge of the molecular pathways that characterize pretransplant lung function has hindered the development of lung-targeted interventions. In this study, we quantified the association between bronchoalveolar lavage (BAL) metatranscriptomes and paired pulmonary function tests performed a median of 1 to 2 weeks before allogeneic HCT in 104 children in The Netherlands. Abnormal pulmonary function was recorded in more than half the cohort, consisted most commonly of restriction and impaired diffusion, and was associated with both all-cause and lung injury–related mortality after HCT. Depletion of commensal supraglottic taxa, such as *Haemophilus*, and enrichment of nasal and skin taxa, such as *Staphylococcus*, in the BAL microbiome were associated with worse measures of lung capacity and gas diffusion. In addition, BAL gene expression signatures of alveolar epithelial activation, epithelial-mesenchymal transition, and down-regulated immunity were associated with impaired lung capacity and diffusion, suggesting a postinjury profibrotic response. Detection of microbial depletion and abnormal epithelial gene expression in BAL enhanced the prognostic utility of pre-HCT pulmonary function tests for the outcome of post-HCT mortality. These findings suggest a potentially actionable connection between microbiome depletion, alveolar injury, and pulmonary fibrosis in the pathogenesis of pre-HCT lung dysfunction.

INTRODUCTION

Allogeneic hematopoietic cell transplantation (HCT) is a life-saving cellular therapy used to cure underlying hematopoietic defects, such as hematologic malignancies, primary immunodeficiencies, bone marrow failure syndromes, and hemoglobinopathies. Although more than 7500 children receive this treatment annually in the United

States and Europe, the success of this therapy is limited by a broad range of post-HCT pulmonary injuries that develop in 10 to 40% of pediatric recipients in response to chemotherapeutic toxicity, infection, and impaired or dysregulated immunity (1–3). Comprehensive assessment of pulmonary health before HCT conditioning is critical to identify high-risk patients who may benefit from modified conditioning chemotherapy regimens, closer surveillance, and risk-stratified treatments aimed at mitigating the development of irreversible pulmonary toxicity (4–6).

In children, evaluation of pulmonary health before HCT is not standardized but may include symptom screening and physical exam, nasopharyngeal swab for common respiratory viruses, pulmonary function testing (PFT) if developmentally able (typically age greater than or equal to 6 years), and targeted high-resolution chest tomography, followed by bronchoalveolar lavage (BAL) in patients with identified or suspected abnormalities (7, 8). Despite the challenges of completing effort-dependent testing in children, when performed with adequate quality, abnormal PFTs are detected in up to 60% of pediatric HCT candidates, predominantly because of restriction and impaired diffusion (9–11), and are associated with the development of post-HCT complications. These include interstitial pneumonitis, pulmonary graft-versus-host disease, and transplant-related mortality (2, 3, 12–16).

To improve outcomes, it remains crucial to identify risk factors for pre-HCT pulmonary dysfunction and elucidate pathobiologic correlates. To date, known risk factors include exposure to alkylating and other pulmotoxic chemotherapy (such as busulfan, bleomycin, cyclophosphamide, carmustine, fludarabine, and lomustine), thoracic radiation (particularly in concert with radiomimetics), thoracic

¹School of Medicine, Department of Pediatrics, Division of Critical Care Medicine, University of California, San Francisco, San Francisco, CA 94143, USA. ²School of Medicine, Department of Pediatrics, Division of Allergy, Immunology, and Bone Marrow Transplantation, University of California, San Francisco, San Francisco, CA 94143, USA. ³University Medical Center Utrecht, Department of Pediatric Stem Cell Transplantation, Utrecht, 3584 CX, Netherlands. ⁴Princess Maxima Center for Pediatric Oncology, Department of Hematopoietic Cell Transplantation, Utrecht 3584 CX, Netherlands. ⁵Department of Pathology, Graduate Program in Experimental Pathology, and Yale Stem Cell Center, Yale University, New Haven, CT 06510, USA. ⁶School of Medicine, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94143, USA. ⁷School of Medicine, Department of Pediatrics, Division of Pulmonology, University of California, San Francisco, San Francisco, CA 94143, USA. ⁸WC Medical College, Department of Pediatrics, Division of Pulmonology, Allergy and Immunology, Cornell University, New York City, NY 10065, USA. ⁹WC Medical College, Department of Pediatrics, Cornell University, New York City, NY 10065, USA. ¹⁰Department of Pediatric Stem Cell Transplantation and Cellular Therapies, Memorial Sloan Kettering Cancer Center, New York City, NY 10065, USA. ¹¹Department of Pediatric Pulmonology, Wilhelmina Children's Hospital, University Medical Centre Utrecht, Utrecht University, Utrecht, 3584 CX, Netherlands. ¹²Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA 94143, USA. ¹³School of Medicine, Department of Pediatrics, University of California, San Francisco, San Francisco, CA 94143, USA. ¹⁴School of Medicine, Cardiovascular Research Institute, Departments of Medicine and Anesthesiology, University of California, San Francisco, San Francisco, CA 94143, USA. ¹⁵Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.

*Corresponding author. Email: matt.zinter@ucsf.edu

†These authors share first authorship.

‡These authors share senior authorship.

surgery, chronic endothelial injury due to hemolysis or transfusion dependence, chronic inflammation, and recurrent pulmonary infections (17–19). Most patients with substantially impaired pre-HCT pulmonary function experience a number of these risk factors, suggesting a multimodal pathogenesis. Despite the need for deeper understanding of the pathogenesis of pre-HCT pulmonary dysfunction, translational studies are limited by scarcity of investigable pulmonary samples from children.

In this study, we aimed to evaluate associations between pre-HCT pulmonary function and measures of pulmonary biology as assessed by metatranscriptomic RNA sequencing of paired BALs. We recently analyzed pre-HCT BAL samples from a cohort of asymptomatic pediatric HCT candidates who all underwent routine pre-HCT BAL in the absence of lower respiratory symptoms as part of an aggressive standard-of-care screening program. We found that pre-HCT pulmonary bacterial depletion, viral enrichment, inflammation, and epithelial activation were associated with the development of post-HCT pulmonary toxicity (20). A subset of these patients were developmentally able to perform contemporaneous PFTs, presenting an opportunity to link the domains of organ function, biology, and microbiology within a cohort of high-risk patients. We hypothesized that the combination of pulmonary microbiome data paired with gene expression data could be associated with abnormal PFTs before HCT, thus providing further evidence for the pathobiologic role of the pulmonary microenvironment in pediatric HCT candidates.

RESULTS

Patient characteristics

Contemporaneous pre-HCT PFTs and BAL fluid were available for 104 patients. PFTs and BAL were typically performed 1 to 2 days apart and 1 to 2 weeks before HCT conditioning. Most patients were between 8 and 19 years of age, of self-reported Northern European ancestry, and preparing to undergo first allogeneic HCT for underlying hematologic malignancy (Table 1). Most patients received myeloablative conditioning chemotherapy, followed by an umbilical cord blood allograft. At 2-year post-HCT follow-up, 33 of 104 patients (31.7%) had died from relapsed disease or transplant toxicity.

Impaired pulmonary function before HCT is common

All patients performed spirometry, and a subset of typically older children performed plethysmography ($n = 78$) and diffusion testing ($n = 63$; Table 2). To determine the interrelatedness of the different PFT measurements, we constructed a correlation matrix, which showed broad grouping of PFTs into four clades pertaining to lung capacity, obstruction, diffusion, and air trapping (fig. S1). Using conventional cutoffs of absolute ratios and percentage deviations from the population mean (% predicted), PFT abnormalities were detected in 80 of 104 patients because of restriction ($n = 35$ of 104), obstruction ($n = 12$ of 104), mixed restriction and obstruction ($n = 1$ of 104), impaired diffusion ($n = 21$ of 63), and air trapping ($n = 44$ of 78). Using Global Lung Index (GLI) cutoffs based on SDs from the population mean (z scores), PFT abnormalities were detected in 54 of 104 patients because of restriction ($n = 29$ of 104), obstruction ($n = 7$ of 104), impaired diffusion ($n = 14$ of 63), and air trapping ($n = 16$ of 78; table S1). As GLI z scores are not calculable for all measured PFTs, we used PFTs expressed as percentage of predicted for subsequent analyses.

Abnormal pulmonary function varied according to age and sex and was associated with HCT outcomes

We tested for clinical characteristics associated with PFT abnormalities and identified that older age was associated with restrictive lung disease and with worse measures of obstruction such as forced expiratory flow (FEF₂₅, FEF₅₀, and FEF₇₅) and peak expiratory flow (PEF) (Fig. 1A). Female sex was associated with worse measures of both lung capacity [including forced vital capacity (FVC) and alveolar volume (V_a)] and diffusion [measured as diffusing capacity for carbon monoxide accounting for V_a (DLCO/V_a)] but better measures of obstruction [such as forced expiratory volume in 1 s adjusted for FVC (FEV₁/FVC); Fig. 1B]. We tested for, but did not identify, an association between HCT indication and pre-HCT PFTs (data file S1).

Previous studies have shown impaired PFTs to be associated with post-HCT mortality (2, 3, 12–16). We confirmed these findings by demonstrating an association between reduced pre-HCT lung capacity and the development of all-cause mortality after HCT (hazard ratio of 1.21 per 10% decline in FVC; 95% confidence interval, 1.03 to 1.43; $P = 0.024$; Fig. 1C). To test whether these poor outcomes were due to acute lung injury specifically, we calculated cumulative incidence with competing risk regression and found that FVC_{%pred} was associated with the development of fatal lung injury in the first 180 days after HCT ($P < 0.001$; fig. S2).

BAL microbiome composition varied with pulmonary function

Our previous studies have associated alterations in the pre-HCT pulmonary microbiome composition with the later development of adverse post-HCT outcomes (20). To determine whether the composition of the pre-HCT pulmonary microbiome is associated with lung function at the time of BAL measurement, we compared microbiome findings to contemporaneously performed PFTs. Microbe RNA sequencing counts are relative in nature and thus were converted to RNA mass abundances using the reference spike-in (21). We first surveyed the composition of BAL microbiomes and found two oppositely correlated clusters of microbes: one containing common supraglottic taxa (such as *Haemophilus*, *Streptococcus*, and *Neisseria*) and a second inversely correlated cluster containing common nasal, skin, and environmental taxa (such as *Staphylococcus* and *Corynebacterium*; fig. S3). We represented these clusters mathematically using principal components analysis (PCA) and found that lower coordinates along principal component 1 (with top contributions from oropharyngeal taxa *Prevotella*, *Veillonella*, *Rothia*, and *Streptococcus*) were associated with worse measures of lung capacity (including FVC; fig. S4 and data file S2).

Next, we used negative binomial generalized linear models (NB-GLMs) with adjustment for false discovery to identify specific taxa associated with each PFT. In models adjusted for patient age and sex, we found that lower BAL quantities of supraglottic taxa (such as *Haemophilus* and *Neisseria*) and greater BAL quantities of skin or environmental taxa (including *Staphylococcus*, *Saccharomyces*, and *Corynebacterium*) were associated with worse measures of lung capacity (FVC), obstruction (FEF₂₅), diffusion impairment (DLCO), and air trapping [residual volume (RV)/total lung capacity (TLC)] (Fig. 2A, fig. S5, and data file S3). For example, progressively lower FVC was associated with progressively less *Haemophilus* RNA and progressively more *Staphylococcus* RNA (Fig. 2, B and C). In addition, greater pulmonary fungal RNA mass was associated with worse measures of lung capacity (FVC) and diffusion (DLCO), and

Table 1. Pediatric allogeneic HCT candidate patient characteristics. Includes $n = 104$ total patients. CMV, cytomegalovirus; HLA, human leukocyte antigen; IQR, interquartile range; N/A, not applicable; PCR, polymerase chain reaction.**Pre-HCT characteristics**

n Age, years, (%)	$4 \geq$ age < 8	24 (23.1)
	$8 \geq$ age < 13	32 (30.8)
	$13 \geq$ age < 20	48 (46.2)
Sex, n (%)	Male	61 (58.3)
	Female	43 (41.4)
Race*, n (%)	Northern European	76 (73.1)
	Multiracial/other	10 (9.6)
	Middle Eastern/Persian/Turkish	7 (6.7)
	African/North African	6 (5.8)
	Asian/Southeast Asian	3 (2.9)
	Eastern European/Russian	2 (1.9)
HCT year, n (%)	2005–2009	23 (22.1)
	2011–2013	40 (38.5)
	2014–2016	41 (39.4)
HCT indication [†] , n (%)	Hematologic malignancy	64 (61.5)
	Bone marrow failure syndrome	17 (16.4)
	Primary immunodeficiency	16 (15.4)
	Metabolic/inborn errors	4 (3.9)
	Autoimmune disease	3 (2.9)
Pre-HCT CMV seropositivity [‡] , n (%)	Positive	56 (53.9)
	Negative	33 (31.7)
	N/A	15 (14.4)
BAL fluid routine microbiology [§] , n (%)	Any abnormality	42 (40.4)
	– Virus detected on PCR	–23 (22.1)
	– Bacterial culture growth	–10 (9.6)
	– Fungal culture growth	–12 (11.5)
	– Galactomannan positivity	–11 (11.2)
Days from PFT to HCT, median (IQR)		11 (8–17)
Days from BAL to HCT, median (IQR)		10.5 (8–14)
Days from PFT to BAL, median (IQR)		1 (–1 to 2)
HCT characteristics and outcomes		
Allograft donor	Mismatched umbilical cord blood	32 (30.8)
	Matched umbilical cord blood	23 (22.1)
	Matched sibling donor	29 (27.9)
	Matched unrelated donor	20 (19.2)
Conditioning [¶]	Myeloablative	97 (93.3)
	Reduced intensity	7 (6.7)
Serotherapy	ATG or alemtuzumab	64 (61.5)
Post-HCT mortality [#]	All-cause mortality	33 (31.7)
	Nonrelapse mortality	21 (20.2)
	– Due to lung injury	–15 (14.4)
	– Not due to lung injury	–6 (5.7)
	Relapse mortality	12 (11.5)

continued on next page

*Race is self-reported. Multiracial/other includes mixed race ($n = 8$), Hispanic ($n = 1$), and Surinamese ($n = 1$). †Malignancy includes acute lymphoid leukemia ($n = 23$), acute myeloid leukemia ($n = 17$), myelodysplastic syndrome ($n = 9$), chronic myeloid leukemia ($n = 4$), and other ($n = 11$). Bone marrow failure syndromes include Fanconi anemia ($n = 11$), severe aplastic anemia ($n = 4$), and other ($n = 2$). Primary immunodeficiency includes combined immunodeficiency ($n = 9$), chronic granulomatous disease ($n = 3$), hemophagocytic lymphohistiocytosis ($n = 2$), and other ($n = 2$). Metabolic/inborn errors include metachromatic leukodystrophy ($n = 2$), adrenoleukodystrophy ($n = 1$), and β -thalassemia ($n = 1$). Autoimmune disease includes systemic lupus erythematosus ($n = 1$), refractory multiple sclerosis ($n = 1$), and systemic juvenile idiopathic arthritis ($n = 1$). ‡CMV serostatus unknown or assumed positive in many recipients of umbilical cord blood allografts. §No organisms were detected with other diagnostic tests such as cytology for *Pneumocystis jirovecii* pneumonia or PCR for mycoplasma or legionella. Some patients had more than one abnormality. ¶Matched sibling donor includes one 8 of 10 HLA match. There were no HLA-mismatched or haploidentical allograft donors. ¶¶Myeloablative conditioning regimens were busulfan and fludarabine (Bu/Flu; $n = 35$); busulfan, fludarabine, and clofarabine (Bu/Flu/Clf; $n = 37$); total body irradiation with etoposide ($n = 11$); fludarabine and cytarabine (Flu/Cy; $n = 10$); and other ($n = 4$). All reduced intensity conditioning regimens were fludarabine based. ††Causes of post-HCT death due to lung injury included invasive fungal infection ($n = 3$), viral infection ($n = 1$), other infection ($n = 1$), alloimmune lung injury including idiopathic pneumonia syndrome ($n = 4$), idiopathic multiorgan failure ($n = 4$), or other ($n = 2$).

Table 2. Pulmonary function tests in pediatric allogeneic HCT candidates. Pulmonary function tests in pediatric allogeneic HCT candidates. Medians with IQR are listed. % of predicted based on Koopman *et al.* age- and sex-adjusted reference sets (61). Z scores based on Global Lung Function Initiative reference values (62). All DLCO measurements are corrected for hemoglobin.

Spirometry* ($n = 104$)	Actual	% of predicted	Z score
Forced vital capacity (FVC; liters)	2.21 (1.50–3.21)	0.87 (0.74–0.96)	–1.07 (–2.28 to –0.35)
Forced expiratory volume in 1 s (FEV ₁ ; liters)	1.94 (1.36–2.77)	0.86 (0.75–0.96)	–1.14 (–2.08 to –0.40)
Forced expiratory volume in 1 s, adjusted (FEV ₁ /FVC)	0.88 (0.84–0.94)	1.00 (0.95–1.06)	–0.17 (–0.83 to 0.92)
Peak expiratory flow (PEF; liters/second)	4.44 (3.12–6.02)	0.85 (0.76–0.97)	–
Forced expiratory flow at 25% (FEF ₂₅ ; liters/s)	3.85 (2.83–5.47)	0.87 (0.72–0.99)	–
Forced expiratory flow at 50% (FEF ₅₀ ; liters/s)	2.58 (1.96–3.59)	0.91 (0.69–1.05)	–
Forced expiratory flow at 75% (FEF ₇₅ ; liters/s)	1.23 (0.86–1.76)	0.86 (0.62–1.06)	–0.40 (–1.28 to 0.32)
Forced expiratory flow 25–75% (FEF _{25–75} ; liters/s)	2.68 (1.73–3.81)	0.98 (0.79–1.15)	–0.10 (–1.00 to 0.55)
Plethysmography† ($n = 78$)			
Total lung capacity (TLC; liters)	3.75 (2.81–4.73)	0.85 (0.74–0.95)	–0.76 (–1.45 to 0.09)
Vital capacity (VC; liters)	2.74 (1.94–3.33)	0.90 (0.75–0.99)	–1.32 (–2.62 to –0.42)
Alveolar volume (V _a ; liters)	3.56 (2.69–4.63)	0.91 (0.80–1.00)	–0.77 (–1.83 to –0.01)
Alveolar volume/total lung capacity (V _a /TLC)	0.95 (0.89–1.01)	0.85 (0.93–0.97)	–
Residual volume (RV; liters)	0.89 (0.71–1.35)	0.99 (0.77–1.30)	0.11 (–0.54 to 0.69)
Residual volume/total lung capacity (RV/TLC)	0.26 (0.21–0.35)	1.06 (0.83–1.44)	0.51 (–0.11 to 1.47)
Diffusion‡ ($n = 63$)			
Diffusion capacity for carbon monoxide (DLCO; mmol/min per kPa)	5.78 (4.35–6.92)	0.78 (0.68–0.95)	–1.65 (–2.53 to –0.34)
Diffusion capacity for carbon monoxide/V _a (DLCO/V _a)	1.50 (1.37–1.73)	0.88 (0.76–1.02)	–0.69 (–1.58 to 0.12)

* $n = 104$ for spirometry. † $n = 78$ for plethysmography. ‡ $n = 63$ for diffusion factor testing.

greater RNA mass of the Actinobacteria phylum, which includes the skin-associated genera *Corynebacterium*, *Micrococcus*, and *Cutibacterium*, was associated with worse measures of lung capacity (FVC) and obstruction (FEF₂₅ and FEF₇₅). In contrast, lower RNA mass of the Fusobacterium phylum was associated with worse measures of lung capacity (V_a) and diffusion (DLCO). We did not find an association between PFTs and total viral RNA mass (data file S2). Some microbiome studies have suggested that aggregate microbiome traits such as alpha diversity, dominance, and richness are associated with the severity of diseases such as chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis (IPF) (22–26). In this cohort, neither alpha diversity, nor richness, nor dominance was associated with PFTs.

BAL gene expression was correlated with pulmonary function

To determine whether impaired pulmonary function was associated with unique gene expression signatures, we first used an overview approach and calculated BAL gene expression enrichment scores for each of the $n = 50$ Hallmark Gene Sets in the Molecular Signatures Database. We found that measures of lung capacity and diffusion were positively associated with the enrichment of pathways relating to immune activation, metabolism, and cell division and were negatively associated with pathways relating to epithelial-mesenchymal transition and cell fate or differentiation (Fig. 3).

We next tested for specific genes whose expression was associated with each PFT. Pathway analysis showed that lower values of both

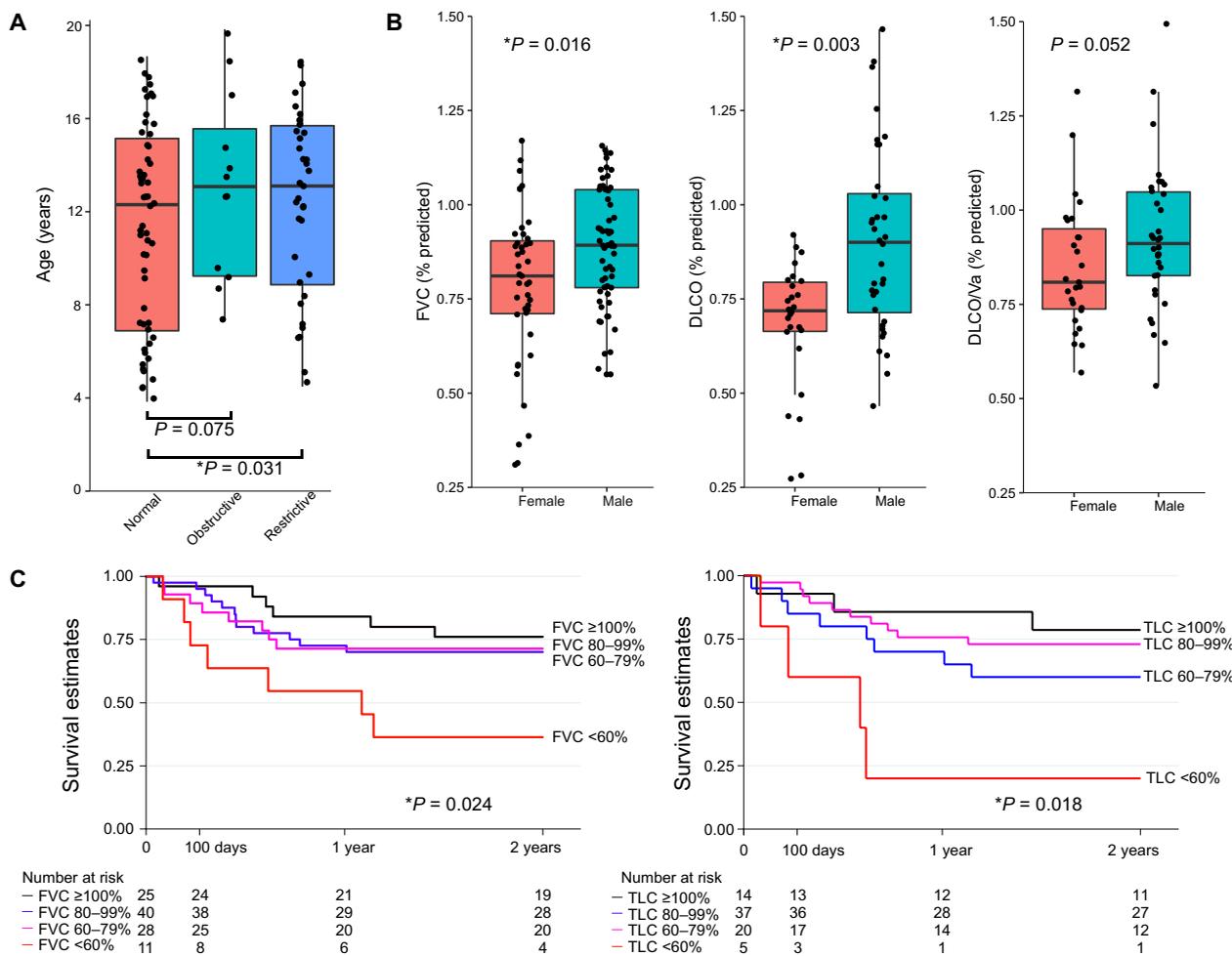


Fig. 1. Association between pre-HCT PFTs, clinical characteristics, and post-HCT outcomes. (A) Box and whisker plots of age distribution are shown for patients with restriction ($n = 35$), obstruction ($n = 12$), or neither ($n = 57$) using conventional PFT categorization. Patients with either restriction or obstruction were older than patients with normal PFTs. Data were analyzed using Dunn's test. (B) Box and whisker plots of PFT distributions are shown according to patient sex (females, $n = 43$; males, $n = 61$). Data were analyzed using Wilcoxon rank sum tests. Box and whisker plots in (A) and (B) depict the median and interquartile range (box) with 1.5 times the interquartile range (whiskers). (C) Kaplan-Meier post-HCT survival estimates are shown according to pre-HCT FVC_{pred} (left) and TLC_{pred} (right). Both FVC_{pred} and TLC_{pred} were associated with post-HCT survival. Data were analyzed by Cox regression.

FVC and DLCO/Va were associated with greater expression of pulmonary epithelial genes, including surfactant genes (such as *SFTPA1/2*, *SFTPC*, and *SFTPD*), keratinization genes (such as *KRTAP3.2* and *KRTAP5.6*), and epithelial sensory and stimulus response genes (such as *OR2A1* and *OR4S2*), and lower expression of innate immunity genes (such as *CXCL8*, *HLA-DRA*, *IL1B*, and *LYZ*; Fig. 4). In contrast, lower FEV1/FVC was associated with greater expression of innate immunity genes and lower expression of alveolar epithelial genes. These findings persisted after adjusting models for age and sex. Few genes ($n = 9$) were associated with a metric of air trapping (RV/TLC; data file S4).

The BAL cell fraction was associated with pulmonary function

Because BAL samples are heterogeneous cell mixtures, we posited that these results could be due to either different proportions of cell types within samples, different expression patterns of each cell type

within samples, or both. Therefore, we used in silico cell deconvolution techniques based on the Human Lung Cell Atlas to impute lung cell proportions in each sample. BAL cell types clustered into three groups: innate immune cells, upper airway cells and lower airway cells, and adaptive immune cells (Fig. 5A). Patients with restriction had greater BAL enrichment scores for type 2 alveolar epithelial cells (AECs) (Fig. 5B). In addition, patients with impaired diffusion had greater BAL enrichment scores for type 2 AECs ($P = 0.011$). Next, using the CIBERSORTx algorithm (27), we estimated the average gene expression profile for each BAL cell type. In patients with restriction, type 1 AECs showed increased transcription of genes related to epithelial-mesenchymal transition, hedgehog signaling, and KRAS signaling. In addition, T and B lymphocytes showed an overall decrease in transcription of genes relating to cell proliferation and inflammation (data file S5). We did not detect differences in innate immune cell or upper airway epithelial cell-specific gene expression among patients with versus without restriction.

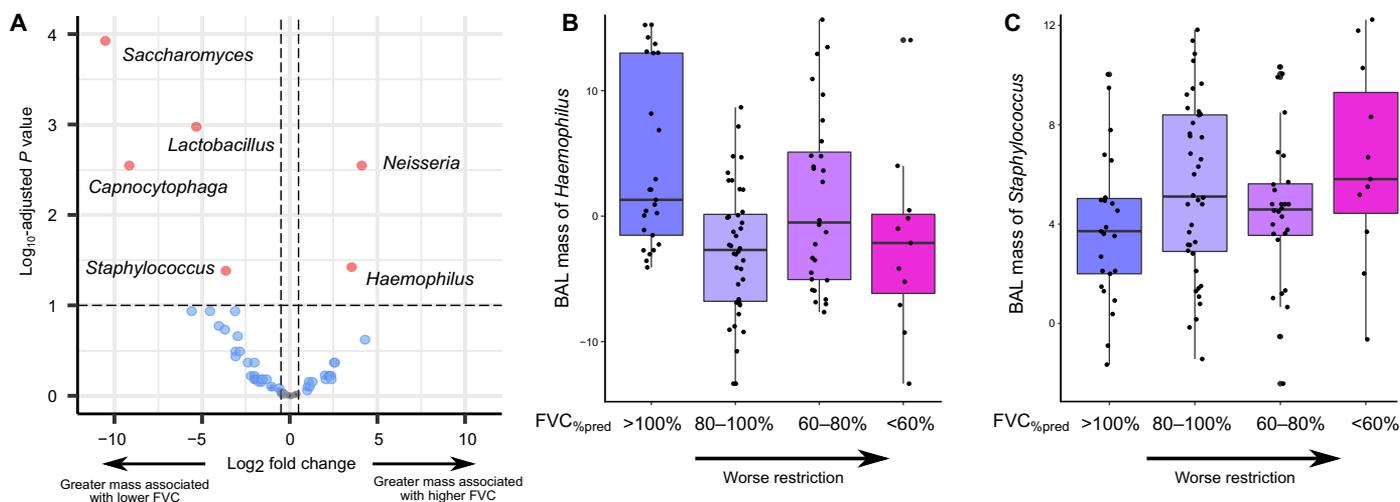


Fig. 2. Association between BAL microbiome and FVC_{%pred}. (A) The volcano plot shows the association between BAL microbe RNA mass and FVC_{%pred}. Associations were tested using negative binomial generalized linear models accounting for patient age and sex covariates. Positive log₂ fold change (x axis) indicates that greater microbial RNA mass is associated with greater FVC_{%pred}; negative log₂ fold change indicates that greater microbial RNA mass is associated with lower FVC_{%pred}. (B and C) Box and whisker plots of variance stabilizing transformed RNA masses of *Haemophilus* (B) and *Staphylococcus* (C) are shown according to category of FVC_{%pred} (>100%, $n = 25$; 80 to 100%, $n = 40$; 60 to 80%, $n = 28$; <60%, $n = 11$). Worse restriction is associated with lower *Haemophilus* and greater *Staphylococcus* RNA mass in BAL fluid. Box and whisker plots in (B) and (C) depict the median and interquartile range (box) with 1.5 times the interquartile range (whiskers).

PFTs and the BAL metatranscriptome are jointly associated with risk of post-HCT mortality

Given previous findings that pre-HCT BAL metatranscriptomes are associated with post-HCT mortality (20), we next sought to determine whether these associations were independent of measured pulmonary function. Using multivariable modeling, we found that lower BAL RNA mass of oropharyngeal taxa (such as *Haemophilus* and *Neisseria*) was associated with post-HCT all-cause mortality independent of FVC, age, and sex. Furthermore, progressively lower BAL RNA mass of oropharyngeal taxa strengthened the association between poor pre-HCT FVC and post-HCT mortality (interaction, $P < 0.05$). In addition, lower BAL expression of genes related to airway epithelial cell integrity, polarity, and junction formation (Hallmark Apical Surface) were also associated with post-HCT mortality independent of FVC, age, and sex (table S2). The addition of BAL *Haemophilus* RNA mass and BAL Hallmark Apical Surface gene set enrichment to a Cox survival model of FVC, age, and sex improved model performance across strata of lung function (likelihood ratio test, $P = 0.002$; Fig. 6).

DISCUSSION

The primary finding of this study is that, in pediatric HCT candidates, pulmonary microbiome and gene expression signatures were associated with pulmonary function and may augment the utility of PFTs in the stratification of post-HCT mortality risk. Specifically, we found that the pulmonary microbiomes of patients with restriction and diffusion impairment were depleted of common supraglottic taxa and instead enriched in skin, nasal, and environmental taxa. In addition, the pulmonary transcriptomes of patients with restriction and diffusion impairment showed increased activation of AECs and a deficit in immune cell signaling consistent with postinjury fibrotic transformation. Furthermore, BAL signatures of commensal microbial depletion and diminished epithelial homeostasis

were associated with post-HCT mortality independent of pre-HCT pulmonary function. Together, these findings suggest a multidimensional connection between lung function, biology, and microbiology in pediatric HCT candidates and possible biologic targets to optimize lung function in children preparing to undergo HCT.

First, our finding that pulmonary microbiome depletion is associated with simultaneously impaired lung function in pediatric HCT candidates supports our previous finding that pre-HCT pulmonary microbiome depletion is associated with post-HCT lung injury (20). Although studies of the pulmonary microbiome in children are rare, investigations of the more readily studied intestinal microbiome have similarly shown that dysbiosis is associated with intestinal disease and poor overall post-HCT outcomes (28). Such landmark studies lend credence to contextualizing an organ-specific microbiome with the harboring organ's function and molecular biology. However, such observations do not prove causality, and therefore whether pulmonary microbial dysbiosis can directly contribute to lung disease or is simply a biomarker of other exposures remains unclear. On the one hand, microbial dysbiosis is linked to antimicrobial exposure, which is more common in patients with underlying diseases such as malignancy; such patients also have confounding reasons for poor lung function, such as greater exposure to pulmotoxic chemotherapy. However, emerging studies suggest that microbial dysbiosis disrupts the control of pathogen overgrowth and contributes to a loss of immune tolerance at the epithelial-parenchymal boundary; therefore, reciprocal causality between lung dysfunction and microbiome disturbances seems likely (22, 29). To that end, we found that the pulmonary RNA mass of taxa that are typically supraglottic commensals was inversely related to the RNA mass of taxa that typically reside in the nose and skin, supporting the theory that maintenance of the respiratory microbiomes may exclude invasion by outside organisms. Future studies directly measuring nasal, supraglottic, and pulmonary microbiomes in each patient may better test whether and how the upper and lower respiratory microbiomes

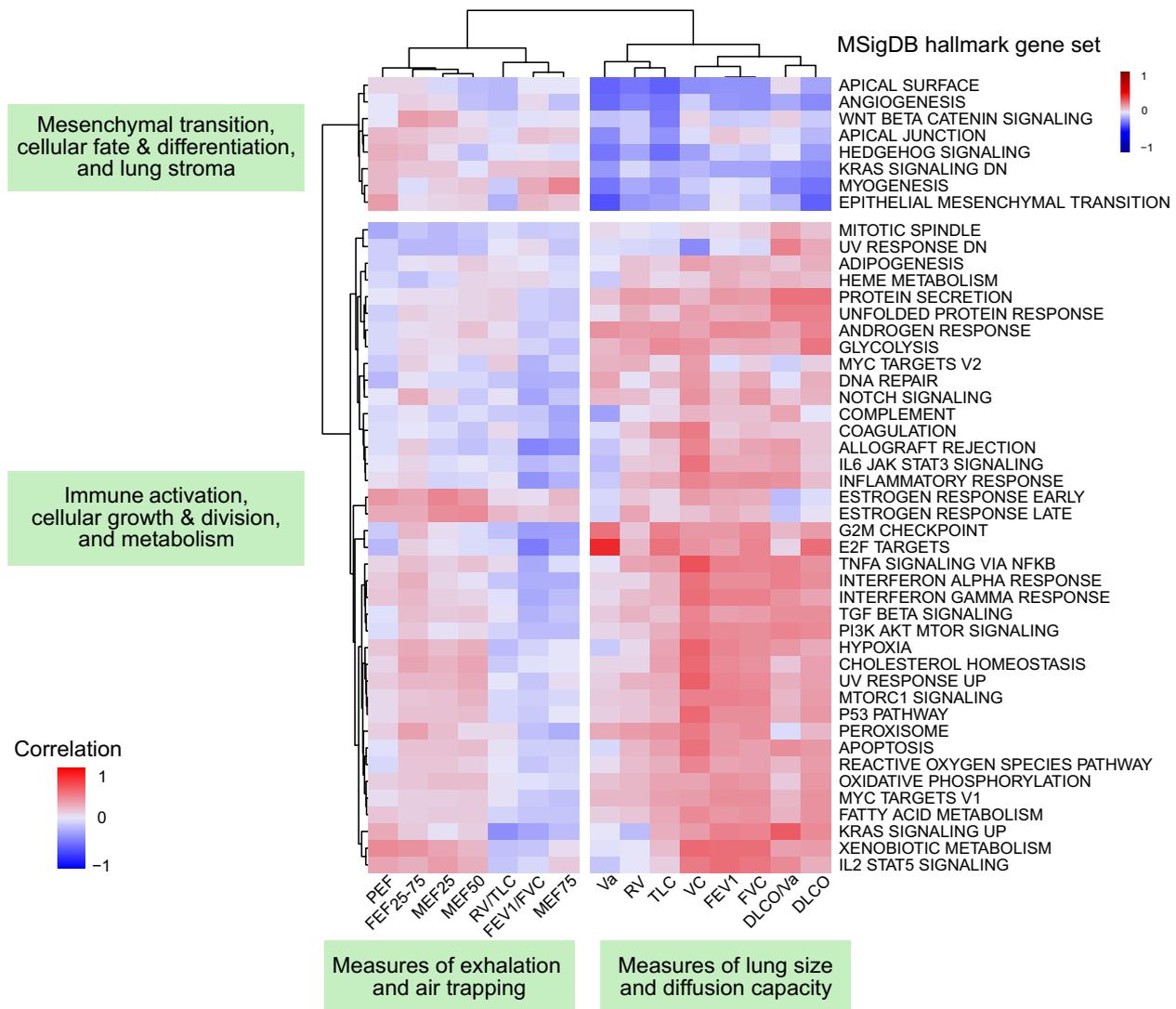


Fig. 3. Association between BAL hallmark gene expression and pulmonary function. Hallmark gene set variance analysis using Poisson count distributions was used to calculate gene set enrichment scores for each patient. Spearman rank-based correlations between gene set enrichment scores (rows) and PFTs (columns) were calculated and plotted (cells) with hierarchical clustering of columns and rows according to Euclidean distance. Dendrogram-based clustering shows that PFT measures of lung size and diffusion were positively associated with a cluster of hallmark gene sets corresponding to immune activation, cellular growth and division, and metabolism and were negatively associated with a cluster of hallmark gene sets corresponding to mesenchymal transition, cellular fate and differentiation, and the lung stroma.

interact and whether the microbiome structures at these upper respiratory sites are indeed microbiologically similar when compared with other populations (30–32). For example, detection of *Staphylococcus* overgrowth in the nose and lungs and *Haemophilus* depletion in the oropharynx and lungs would provide more direct evidence for the assertion that the supraglottic microbiome protects the lungs from nasal, skin, and environmental microbes. In addition, strategies to characterize gut-lung cross-talk may further elucidate factors influencing the pulmonary microbiome (33, 34). Ultimately, strategies to modulate the lung microbiome to improve pulmonary health remain in development and include investigation of both ingested and inhaled probiotics and other microbial metabolites (35, 36).

A second finding of this work is the association between altered alveolar epithelial gene expression, restriction, and impaired diffusion.

By combining differential gene expression with cell type imputation, we identified a greater proportion of type 2 AECs in patients with restriction and diffusion; furthermore, cell type-specific gene expression suggested an increased transcriptional state of these cells, associated with growth factors and increased kinase expression. The increased expression of surfactant genes and overall AEC activity are consistent with response to alveolar injury and have been demonstrated in numerous cohorts of lung-injured patients (37, 38). By combining hallmark gene set enrichment with these findings, we found that pathways related to the lung stroma were up-regulated, including angiogenesis, myogenesis, epithelial-mesenchymal transition, and cell fate (Wnt/ β -catenin signaling and Hedgehog signaling). Activation of stromal pathways in the lung has been associated with a profibrotic signaling state and has been identified in cohorts of patients with IPF, thus providing a plausible

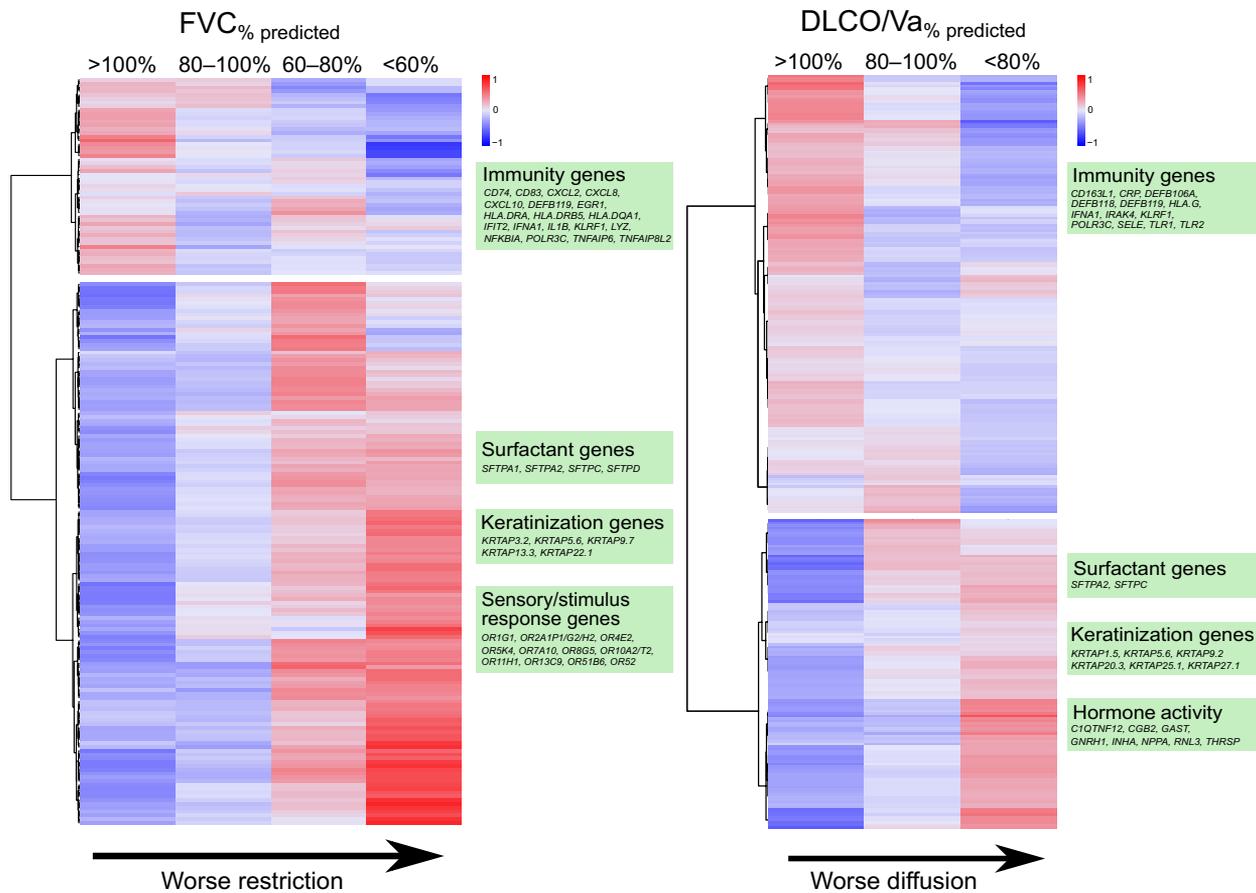


Fig. 4. Association between BAL gene expression and PFT. Heatmaps of differential gene expression are plotted according to pulmonary restriction (left) and impaired diffusion (right). Negative binomial models were used to identify differentially expressed genes (rows) across the continuum of $FVC_{\%pred}$ and $DLCO/Va_{\%pred}$ (columns). Cell shading indicates mean gene expression for each PFT subgroup after variance-stabilizing transformation, centering, and scaling. Dendrogram-based clustering shows that declining $FVC_{\%pred}$ was associated with a decrease in expression of immune-related genes and an increase in expression of airway epithelial genes relating to surfactant, keratinization, and olfactory responses. Declining $DLCO/Va_{\%pred}$ was also associated with a decrease in expression of immune-related genes and an increase in expression of airway epithelial genes and hormone response genes.

mechanistic link between these gene expression findings and the observed impairment in PFTs (39, 40). As lung-resident mesenchymal stem cells undergoing myofibroblastic differentiation strongly express genes in the Wnt/ β -catenin and Hedgehog signaling pathways, antifibrotic agents targeting this pathway, such as pirfenidone, remain of utmost interest (41, 42).

A third finding of this work is that patients with pulmonary restriction and impaired diffusion displayed an overall deficit in lower airway immune signaling. In addition to lower overall expression of immune genes among patients with restriction, cell type deconvolution inferred a less active transcriptional state for lymphocytes specifically. It seems counterintuitive that patients with poor PFTs would have a relative deficit in pulmonary immune activity, especially given our previous findings that pre-HCT pulmonary inflammation was associated with post-HCT lung injury. There are three potential explanations. First, the latter finding of pulmonary inflammation was tied to pre-HCT viral infection, which was more prevalent in children younger than 4 years. As younger children are not typically able to perform PFTs, they were excluded from this present study. Second, although pulmonary inflammation is a hallmark of acute lung injury including chemoradiation-induced lung

injury, it typically subsides after the initial insult and is followed by a profound anti-inflammatory and profibrotic phase (43), which more likely represents the phase of injury for the patients in this study. Third, patients with impaired lower airway immunity may be at greater risk for infections and associated sequelae such as postinfectious fibrosis. Ultimately, the relevance of impaired pulmonary immune environment in pediatric HCT candidates with poor lung function needs to be explored with additional approaches in the future.

A fourth and unanticipated finding of this work was the identification of substantial sex differences in pulmonary function. Although both male and female study participants had worse lung function than age- and sex-matched controls, females had profoundly worse measures of both restriction and impaired diffusion, and males had slightly worse measures of obstruction. Long-term follow-up studies of pediatric oncology and HCT patients have also demonstrated associations between female sex, restriction, and impaired diffusion, as well as between male sex and obstruction; this study adds to the field by identifying these disparities before HCT. A likely culprit for these toxicities are underlying differences in chemotherapy pharmacokinetics and pharmacodynamics, which merit exploration in future studies (44–46).

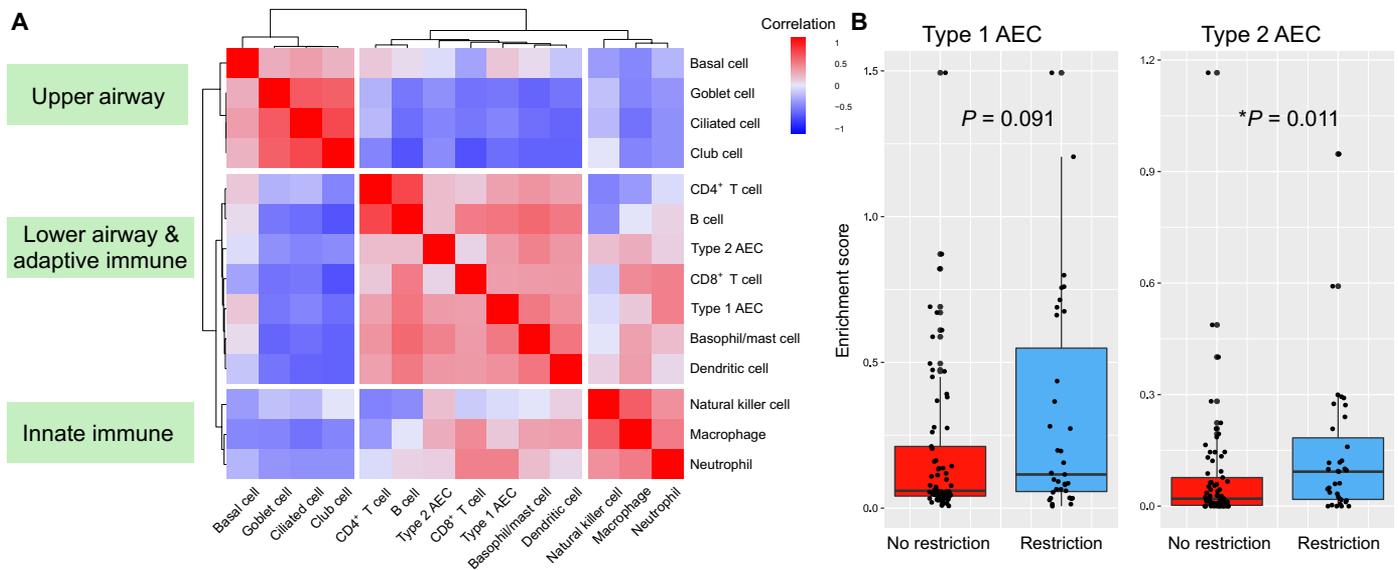


Fig. 5. BAL cell type differential. Single-cell RNA sequencing data from The Human Lung Cell Atlas was used to calculate cell type-specific marker genes and impute cell type fractions and enrichment scores (CIBERSORTx). **(A)** Highly correlated cell types were identified by calculating Spearman rank-based correlation of cell type fractions followed by hierarchical clustering of correlation distances. **(B)** Cell type enrichment scores were tested for association with pulmonary restriction (no restriction, $n = 69$; restriction, $n = 35$) using Kruskal-Wallis testing.

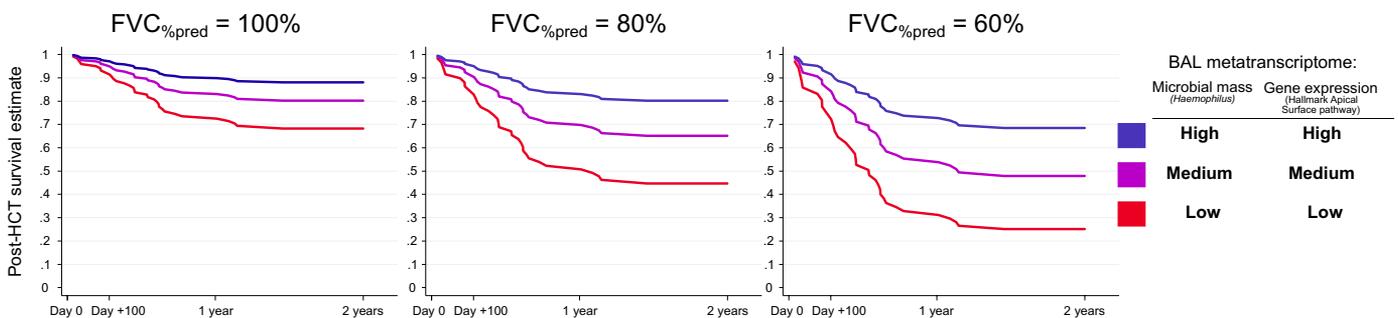


Fig. 6. Survival estimates by pre-HCT BAL metatranscriptome and pre-HCT FVC. A multivariable Cox regression model for the outcome of post-HCT all-cause mortality was fit for all patients ($n = 104$) using pre-HCT FVC_{pred} , pre-HCT BAL RNA mass of *Haemophilus* (variance-stabilizing transformed), and pre-HCT BAL enrichment score of MSigDB Hallmark Apical Surface gene set with covariate adjustment for patient age and sex. Model-based mortality estimates are plotted according to pre-HCT FVC (100% versus 80% versus 60%) and high, medium, and low *Haemophilus* RNA mass and apical surface gene expression. High, medium, and low refer to the 75th, 50th, and 25th percentile within the cohort, respectively.

Our study has several strengths. First, the analysis of contemporaneously measured PFTs and BAL is unique among pediatric cohorts and allows multidimensional insights into lung function, biology, and microbiology. Second, we provide PFT measurements using both percent of predicted and GLI z scores according to updated American Thoracic Society/European Respiratory Society recommendations. Third, we analyzed BAL using a paired assessment of microbiome and gene expression.

Several limitations merit comment. First, as with all observational human studies, we are unable to prove mechanism or causation among the factors measured and described herein. Second, we were unable to ascertain all possible risk factors that may have influenced the pulmonary microenvironment, including history of pulmonary infections, prior antimicrobial exposure, and detailed chemotherapeutic regimens. Third, BAL samples were not paired with supraglottic and nasopharyngeal samples, and therefore, assumptions about the

possible origin of BAL taxa described as typically supraglottic or typically nasal or environmental are based on published literature rather than direct observation. Fourth, the identification of microbial RNA does not directly imply viable, intact organisms in the lower airways; future studies leveraging functional microbial genomics and metabolomics may provide deeper insight into the activity states of detected organisms (47, 48). Fifth, we did not have documentation of the fractional return of each patient's lavage. Low fractional return is associated with both restrictive and obstructive lung disease, and yield below 30% can indicate undersample of the distal airways. Of note, undersampling of the distal airways would be associated with greater microbial recovery and lower AEC recovery, which is the opposite of what we found in our patients with pulmonary restriction (49–54). Last, younger children were necessarily excluded because of their inability to perform PFTs. However, given our findings associating PFTs with lower respiratory biology,

this study raises the possibility that BAL RNA sequencing could be a useful surrogate to detect profibrotic states in those unable to perform PFTs. Greater correlation between biologic measures, standard spirometry, and newer PFT approaches such as forced impulse oscillometry and multiple breath washout may yield additional insights into children unable to perform standard PFTs (55–58).

In conclusion, in this study, we found that among children preparing to undergo allogeneic HCT, PFT abnormalities were common and consisted mostly of restriction and impaired diffusion. Abnormalities were associated with pulmonary microbiome depletion, profibrotic signaling, and post-HCT mortality. These findings suggest a potentially actionable connection between microbiome depletion, alveolar injury, and pulmonary fibrosis in the pathogenesis of HCT-related lung dysfunction.

MATERIALS AND METHODS

Study design

As previously reported, pediatric allogeneic HCT candidates up to 19 years of age with malignant and nonmalignant diseases underwent protocolized evaluation of pre-HCT pulmonary health between 2005 and 2016 in Utrecht, The Netherlands (11). Patients were not randomized, and investigators were not blinded. This study included the subset who completed both pre-HCT PFTs and pre-HCT BAL before the start of conditioning chemotherapy ($n = 104$). Informed consent was obtained with Institutional Review Board approval (#05/143 and #11/063). Details of approach to peri-HCT clinical care are described in the Supplementary Materials and Methods. Because of the exploratory nature of this study, a priori sample size estimates were not performed for the analyses described herein; individual-level data for analyses with groups consisting of fewer than 20 patients are provided in data file S6.

PFT measurements

PFTs included spirometry, whole-body plethysmography, and carbon monoxide diffusion capacity and were performed in developmentally capable children 4 years and older according to established criteria (59, 60). Prebronchodilator PFT values of adequate performance quality, as assessed by each patient's clinical pulmonologist, were normalized for age, height, sex, and race and expressed as (i) a percentage deviation from the mean (% predicted) according to healthy Dutch children and (ii) an SD from the mean (z score) according to the GLI (61, 62). Restriction, obstruction, mixed restriction and obstruction, diffusion impairment, and air trapping were identified first using conventional PFT cutoffs and were compared with GLI-recommended cutoffs wherein PFTs below the 5th percentile (z score < -1.64) were considered below the lower limit of normal (Supplementary Materials and Methods) (59, 62).

BAL measurements

BAL was performed in all pediatric HCT candidates as standard-of-care routine clinical practice to identify potential pulmonary pathogens and reduce the risk of post-HCT lung injury. BAL was performed in the absence of lower respiratory tract symptoms and at the time of another clinically indicated procedure (such as bone marrow aspirate and central line placement) (19, 63). Cryopreserved aliquots of BAL underwent metatranscriptomic RNA sequencing as previously described (20). Briefly, samples underwent mechanical homogenization followed by RNA extraction and purification (64, 65).

Sample RNA was combined with control spike-in RNA and underwent reverse transcription, library preparation, and 125-nucleotide paired-end sequencing on a NovaSeq 6000 instrument to a target depth of 40 million read pairs per sample (Supplementary Materials and Methods). Resultant .fastq files underwent quality filtration followed by parallel alignment to both human and microbial reference genomes using the IDseq pipeline (66). Microbe count data were adjusted for laboratory contamination, and estimated microbe RNA mass was calculated using the reference External RNA Controls Consortium (ERCC) spike-in (Supplementary Materials Methods) (21).

Statistical analysis

Associations between PFTs and patient characteristics were tested using Kruskal-Wallis tests and Spearman correlations for categorical and continuous traits, respectively. Associations between pre-HCT PFTs and post-HCT outcomes were tested with Kaplan-Meier estimates and Cox proportional hazards models for all-cause mortality and cumulative incidence function with competing risk regression for fatal lung injury. The pulmonary microbiome was described using taxa correlation matrices and PCA. Associations between PFTs and microbial RNA masses were tested using NB-GLM with adjustment for false discovery in the R statistical environment (edgeR) (67). Gene expression was characterized by calculating enrichment scores for the Hallmark Gene Sets in the Molecular Signatures Database (MSigDB) (68). Associations between PFTs and differentially expressed genes were tested using NB-GLM and characterized using pathway analysis (enrichR) (69). BAL cell type proportions and cell type-specific gene expression profiles were imputed using CIBERSORTx and the Human Lung Cell Atlas (Supplementary Materials and Methods and fig. S6) (27, 70).

SUPPLEMENTARY MATERIALS

www.science.org/doi/10.1126/scitranslmed.abm8646

Materials and Methods

Figs. S1 to S6

Tables S1 and S2

MDAR Reproducibility Checklist

Data files S1 to S6

References (71–73)

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. A. D'Souza, C. Fretham, S. J. Lee, M. Aurora, J. Brunner, S. Chhabra, S. Devine, M. Eapen, M. Hamadani, P. Hari, M. C. Pasquini, R. A. Phelan, M. L. Riches, J. D. Rizzo, W. Saber, B. E. Shaw, S. R. Spellman, P. Steinert, D. J. Weisdorf, M. M. Horowitz, Current use of and trends in hematopoietic cell transplantation in the United States. *Biol. Blood Marrow Transplant.* **26**, e177–e182 (2020).
2. L. Broglie, C. Fretham, A. Al-Seraihy, B. George, J. Kurtzberg, A. Loren, M. MacMillan, C. Martinez, S. M. Davies, M. C. Pasquini, Pulmonary complications in pediatric and adolescent patients following allogeneic hematopoietic cell transplantation. *Biol. Blood Marrow Transplant.* **25**, 2024–2030 (2019).
3. Z. Kaya, D. J. Weiner, D. Yilmaz, J. Rowan, R. K. Goyal, Lung function, pulmonary complications, and mortality after allogeneic blood and marrow transplantation in children. *Biol. Blood Marrow Transplant.* **15**, 817–826 (2009).
4. R. F. Tamburro, K. R. Cooke, S. M. Davies, S. Goldfarb, J. S. Hagood, A. Srinivasan, M. E. Steiner, D. Stokes, N. DiFronzo, N. El-Kassar, N. Shelburne, A. Natarajan; Pulmonary Complications of Pediatric Hematopoietic Stem Cell Transplantation Workshop Participants, Pulmonary complications of pediatric hematopoietic stem cell transplantation (HCT): An NIH workshop summary. *Ann. Am. Thorac. Soc.* **18**, 381–394 (2021).
5. J. Duque-Afonso, G. Ihorst, M. Waterhouse, R. Zeiser, R. Wäsch, H. Bertz, M. Yücel, T. Köhler, J. Müller-Quernheim, R. Marks, J. Finke, Comparison of reduced-toxicity conditioning protocols using fludarabine, melphalan combined with thiopeta or carmustine in allogeneic hematopoietic cell transplantation. *Bone Marrow Transplant.* **56**, 110–120 (2021).

6. T. Schechter, A. Naqvi, S. Weitzman, Risk for complications in patients with hemophagocytic lymphohistiocytosis who undergo hematopoietic stem cell transplantation: Myeloablative versus reduced-intensity conditioning regimens. *Expert Rev. Clin. Immunol.* **10**, 1101–1106 (2014).
7. G.-S. Cheng, Pulmonary function and pretransplant evaluation of the hematopoietic cell transplant candidate. *Clin. Chest Med.* **38**, 307–316 (2017).
8. J. W. Chien, D. K. Madtes, J. G. Clark, Pulmonary function testing prior to hematopoietic stem cell transplantation. *Bone Marrow Transplant.* **35**, 429–435 (2005).
9. J. Wieringa, K. W. van Kralingen, J. K. Sont, D. Bresters, Pulmonary function impairment in children following hematopoietic stem cell transplantation. *Pediatr. Blood Cancer* **45**, 318–323 (2005).
10. A. R. Smith, N. S. Majhail, M. L. MacMillan, T. E. DeFor, S. Jodele, L. E. Lehmann, R. Krance, S. M. Davies, Hematopoietic cell transplantation comorbidity index predicts transplantation outcomes in pediatric patients. *Blood* **117**, 2728–2734 (2011).
11. A. B. Versluys, K. van der Ent, J. J. Boelens, T. Wolfs, P. de Jong, M. B. Bierings, High diagnostic yield of dedicated pulmonary screening before hematopoietic cell transplantation in children. *Biol. Blood Marrow Transplant.* **21**, 1622–1626 (2015).
12. A. Srinivasan, S. Srinivasan, S. Sunthakar, A. Sunkara, G. Kang, D. C. Stokes, W. Leung, Pre-hematopoietic stem cell transplant lung function and pulmonary complications in children. *Ann. Am. Thorac. Soc.* **11**, 1576–1585 (2014).
13. J. P. Ginsberg, R. Aplenc, J. McDonough, J. Bethel, J. Doyle, D. J. Weiner, Pre-transplant lung function is predictive of survival following pediatric bone marrow transplantation. *Pediatr. Blood Cancer* **54**, 454–460 (2010).
14. T. C. Quigg, Y.-J. Kim, W. S. Goebel, P. R. Haut, Lung function before and after pediatric allogeneic hematopoietic stem cell transplantation: A predictive role for DLCOa/VA. *J. Pediatr. Hematol. Oncol.* **34**, 304–309 (2012).
15. S. Herasevich, R. D. Frank, H. Bo, H. Alkhatieb, W. J. Hogan, O. Gajic, H. Yadav, Pretransplant risk factors can predict development of acute respiratory distress syndrome after hematopoietic stem cell transplantation. *Ann. Am. Thorac. Soc.* **18**, 1004–1012 (2021).
16. H. J. Lee, K. Kim, S. K. Kim, J. W. Lee, J.-S. Yoon, N.-G. Chung, C. Bin, Hb-adjusted DLCO with GLI reference predicts long-term survival after HSCT in children. *Bone Marrow Transplant.* **56**, 1929–1936 (2021).
17. T.-T. Huang, M. M. Hudson, D. C. Stokes, M. J. Krasin, S. L. Spunt, K. K. Ness, Pulmonary outcomes in survivors of childhood cancer: A systematic review. *Chest* **140**, 881–901 (2011).
18. S. Kharbanda, A. Panoskaltis-Mortari, I. Y. Haddad, B. R. Blazar, P. J. Orchard, D. N. Cornfield, S. S. Grewal, C. Peters, W. E. Regelman, C. E. Milla, K. S. Baker, Inflammatory cytokines and the development of pulmonary complications after allogeneic hematopoietic cell transplantation in patients with inherited metabolic storage disorders. *Biol. Blood Marrow Transplant.* **12**, 430–437 (2006).
19. B. Versluys, M. Bierings, J. L. Murk, T. Wolfs, C. Lindemans, K. Vd Ent, J. J. Boelens, Infection with a respiratory virus before hematopoietic cell transplantation is associated with alloimmune-mediated lung syndromes. *J. Allergy Clin. Immunol.* **141**, 697–703.e8 (2018).
20. M. S. Zinter, C. A. Lindemans, B. A. Versluys, M. Y. Mayday, S. Sunshine, G. Reyes, M. Sirota, A. Sapru, M. A. Matthay, S. Kharbanda, C. C. Dvorak, J. J. Boelens, J. L. DeRisi, The pulmonary metatranscriptome prior to pediatric HCT identifies post-HCT lung injury. *Blood* **137**, 1679–1689 (2021).
21. M. S. Zinter, M. Y. Mayday, K. K. Ryckman, L. L. Jelliffe-Pawlowski, J. L. DeRisi, Towards precision quantification of contamination in metagenomic sequencing experiments. *Microbiome* **7**, 62 (2019).
22. D. N. O'Dwyer, X. Zhou, C. A. Wilke, M. Xia, N. R. Falkowski, K. C. Norman, K. B. Arnold, G. B. Huffnagle, S. Murray, J. R. Erb-Downward, G. A. Yanik, B. B. Moore, R. P. Dickson, Lung dysbiosis, inflammation, and injury in hematopoietic cell transplantation. *Am. J. Respir. Crit. Care Med.* **198**, 1312–1321 (2018).
23. D. N. O'Dwyer, S. L. Ashley, S. J. Gurczynski, M. Xia, C. Wilke, N. R. Falkowski, K. C. Norman, K. B. Arnold, G. B. Huffnagle, M. L. Salisbury, M. K. Han, K. R. Flaherty, E. S. White, F. J. Martinez, J. R. Erb-Downward, S. Murray, B. B. Moore, R. P. Dickson, Lung microbiota contribute to pulmonary inflammation and disease progression in pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1127–1138 (2019).
24. G. G. Einarsson, D. M. Comer, L. McIlreavey, J. Parkhill, M. Ennis, M. M. Tunney, J. S. Elborn, Community dynamics and the lower airway microbiota in stable chronic obstructive pulmonary disease, smokers and healthy non-smokers. *Thorax* **71**, 795–803 (2016).
25. A. A. Pragman, H. B. Kim, C. S. Reilly, C. Wendt, R. E. Isaacson, The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLOS ONE* **7**, e47305 (2012).
26. S. Gupta, M. Shariff, G. Chaturvedi, A. Sharma, N. Goel, M. Yadav, M. S. Mortensen, S. J. Sørensen, M. Mukerji, N. S. Chauhan, Comparative analysis of the alveolar microbiome in COPD, EOPD, Sarcoidosis, and ILD patients to identify respiratory illnesses specific microbial signatures. *Sci. Rep.* **11**, 3963 (2021).
27. A. M. Newman, C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, M. S. Khodadoust, M. S. Esfahani, B. A. Luca, D. Steiner, M. Diehn, A. A. Alizadeh, Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
28. J. U. Peled, A. L. C. Gomes, S. M. Devlin, E. R. Littmann, Y. Taur, A. D. Sung, D. Weber, D. Hashimoto, A. E. Slingerland, J. B. Slingerland, M. Maloy, A. G. Clurman, C. K. Stein-Thoeringer, K. A. Markey, M. D. Docampo, M. B. da Silva, N. Khan, A. Gessner, J. A. Messina, K. Romero, M. V. Lew, A. Bush, L. Bohannon, D. G. Brereton, E. Fontana, L. A. Amoretti, R. J. Wright, G. K. Armijo, Y. Shono, M. Sanchez-Escamilla, N. C. Flores, A. A. Tomas, R. J. Lin, L. Y. S. Segundo, G. L. Shah, C. Cho, M. Scordo, I. Politikos, K. Hayasaka, Y. Hasegawa, B. Gyurkocza, D. M. Ponce, J. N. Barker, M. A. Perales, S. A. Giral, R. R. Jenq, T. Teshima, N. J. Chao, E. Holler, J. B. Xavier, E. G. Pamer, M. R. M. van den Brink, Microbiota as predictor of mortality in allogeneic hematopoietic-cell transplantation. *N. Engl. J. Med.* **382**, 822–834 (2020).
29. N. A. Abreu, N. A. Nagalingam, Y. Song, F. C. Roediger, S. D. Pletcher, A. N. Goldberg, S. V. Lynch, Sinus microbiome diversity depletion and *Corynebacterium tuberculoearicum* enrichment mediates rhinosinusitis. *Sci. Transl. Med.* **4**, 151ra124 (2012).
30. L. N. Segal, A. V. Alekseyenko, J. C. Clemente, R. Kulkarni, B. Wu, Z. Gao, H. Chen, K. I. Berger, R. M. Goldring, W. N. Rom, M. J. Blaser, M. D. Weiden, Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome* **1**, 19–19 (2013).
31. R. P. Dickson, J. R. Erb-Downward, C. M. Freeman, L. McCloskey, N. R. Falkowski, G. B. Huffnagle, J. L. Curtis, Bacterial topography of the healthy human lower respiratory tract. *MBio* **8**, 10.1128/mBio.02287-16, (2017).
32. R. P. Dickson, J. R. Erb-Downward, F. J. Martinez, G. B. Huffnagle, The microbiome and the respiratory tract. *Annu. Rev. Physiol.* **78**, 481–504 (2016).
33. R. Trivedi, K. Barve, Gut microbiome a promising target for management of respiratory diseases. *Biochem. J.* **477**, 2679–2696 (2020).
34. T. J. Schuijt, J. M. Lankelma, B. P. Scicluna, F. de Sousa e Melo, J. J. T. H. Roelofs, J. D. de Boer, A. J. Hoogendijk, R. de Beer, A. de Vos, C. Belzer, W. M. de Vos, T. van der Poll, W. J. Wiersinga, The gut microbiota plays a protective role in the host defence against pneumococcal pneumonia. *Gut* **65**, 575–583 (2016).
35. H.-S. Kim, K. Arellano, H. Park, S. D. Todorov, B. Kim, H. Kang, Y. J. Park, D. H. Suh, E. S. Jung, Y. Ji, W. H. Holzappel, Assessment of the safety and anti-inflammatory effects of three *Bacillus* strains in the respiratory tract. *Environ. Microbiol.* **23**, 3077–3098 (2021).
36. A. Nieto, A. Mazón, M. Nieto, R. Calderón, S. Calaforra, B. Selva, S. Uixera, M. J. Palao, P. Brandi, L. Conejero, P. Saz-Leal, C. Fernández-Pérez, D. Sancho, J. L. Subiza, M. Casanovas, Bacterial mucosal immunotherapy with MV130 prevents recurrent wheezing in children: A randomized, double-blind, placebo-controlled clinical trial. *Am. J. Respir. Crit. Care Med.* **204**, 462–472 (2021).
37. O. Garcia, M. J. Hiatt, A. Lundin, J. Lee, R. Reddy, S. Navarro, A. Kikuchi, B. Driscoll, Targeted type 2 alveolar cell depletion: a dynamic functional model for lung injury repair. *Am. J. Respir. Cell. Mol. Biol.* **54**, 319–330 (2016).
38. M. K. Dahmer, H. Flori, A. Sapru, J. Kohne, H. M. Weeks, M. A. Q. Curley, M. A. Matthay, M. W. Quasney; BALI and RESTORE Study Investigators and Pediatric Acute Lung Injury and Sepsis Investigators (PALISI) Network, Surfactant protein D is associated with severe pediatric ARDS, prolonged ventilation, and death in children with acute respiratory failure. *Chest* **158**, 1027–1035 (2020).
39. J. H. W. Distler, A.-H. Györfi, M. Ramanujam, M. L. Whitfield, M. Königshoff, R. Lafyatis, Shared and distinct mechanisms of fibrosis. *Nat. Rev. Rheumatol.* **15**, 705–730 (2019).
40. L. M. Crosby, C. M. Waters, Epithelial repair mechanisms in the lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **298**, L715–L731 (2010).
41. B. Hu, J. Liu, Z. Wu, T. Liu, M. R. Ullenbruch, L. Ding, C. A. Henke, P. B. Bitterman, S. H. Phan, Reemergence of hedgehog mediates epithelial-mesenchymal crosstalk in pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **52**, 418–428 (2015).
42. M. Didiasova, R. Singh, J. Wilhelm, G. Kwapiszewska, L. Wujak, D. Zakrzewicz, L. Schaefer, P. Markart, W. Seeger, M. Lauth, M. Wygrecka, Pirfenidone exerts antifibrotic effects through inhibition of GLI transcription factors. *FASEB J.* **31**, 1916–1928 (2017).
43. R. Malaviya, H. M. Kipen, R. Businaro, J. D. Laskin, D. L. Laskin, Pulmonary toxicants and fibrosis: Innate and adaptive immune mechanisms. *Toxicol. Appl. Pharmacol.* **409**, 115272 (2020).
44. S. Haupt, F. Caramia, S. L. Klein, J. B. Rubin, Y. Haupt, Sex disparities matter in cancer development and therapy. *Nat. Rev. Cancer* **21**, 393–407 (2021).
45. A. D. Wagner, A. Grothey, T. Andre, J. G. Dixon, N. Wolmark, D. G. Haller, C. J. Allegra, A. de Gramont, E. VanCutsem, S. R. Alberts, T. J. George, M. J. O'Connell, C. Twelves, J. Taieb, L. B. Saltz, C. D. Blanke, E. Francini, R. Kerr, G. Yothers, J. F. Seitz, S. Marsoni, R. M. Goldberg, Q. Shi, Sex and adverse events of adjuvant chemotherapy in colon cancer: An analysis of 34 640 patients in the ACCENT database. *J. Natl. Cancer Inst.* **113**, 400–407 (2021).
46. M. Diefenhardt, E. B. Ludmir, R.-D. Hofheinz, M. Ghadimi, B. D. Minsky, C. Rödel, E. Fokas, Association of sex with toxic effects, treatment adherence, and oncologic outcomes

- in the CAO/ARO/AIO-94 and CAO/ARO/AIO-04 Phase 3 randomized clinical trials of rectal cancer. *JAMA Oncol.* **6**, 294–296 (2020).
47. I. Sulaiman, B. G. Wu, Y. Li, J.-C. Tsay, M. Sauthoff, A. S. Scott, K. Ji, S. B. Koralov, M. Weiden, J. C. Clemente, D. Jones, Y. J. Huang, K. A. Stringer, L. Zhang, A. Geber, S. Banakis, L. Tipton, E. Ghedin, L. N. Segal, Functional lower airways genomic profiling of the microbiome to capture active microbial metabolism. *Eur. Respir. J.* **58**, 2003434 (2021).
 48. Z. Wang, Y. Yang, Z. Yan, H. Liu, B. Chen, Z. Liang, F. Wang, B. E. Miller, R. Tal-Singer, X. Yi, J. Li, M. R. Stampfli, H. Zhou, C. E. Brightling, J. R. Brown, M. Wu, R. Chen, W. Shu, Multi-omic meta-analysis identifies functional signatures of airway microbiome in chronic obstructive pulmonary disease. *ISME J.* **14**, 2748–2765 (2020).
 49. K. C. Meyer, G. Raghu, R. P. Baughman, K. K. Brown, U. Costabel, R. M. du Bois, M. Drent, P. L. Haslam, D. S. Kim, S. Nagai, P. Rottoli, C. Saltini, M. Selman, C. Strange, B. Wood; American Thoracic Society Committee on BAL in Interstitial Lung Disease, An official american thoracic society clinical practice guideline: The clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. *Am. J. Respir. Crit. Care Med.* **185**, 1004–1014 (2012).
 50. M. Heron, J. C. Grutters, K. M. ten Dam-Molenkamp, D. Hijdra, A. van Heugten-Roeling, A. M. E. Claessen, H. J. T. Ruven, J. M. M. van den Bosch, H. van Velzen-Blad, Bronchoalveolar lavage cell pattern from healthy human lung. *Clin. Exp. Immunol.* **167**, 523–531 (2012).
 51. K. Shikano, M. Abe, Y. Shiko, K. Tsushima, K. Yoshioka, T. Ishiwata, T. Kawasaki, J. Ikari, J. Terada, Y. Kawasaki, K. Tatsumi, What are the factors affecting the recovery rate of bronchoalveolar lavage fluid? *Clin. Respir. J.* **16**, 142–151 (2021).
 52. J. Schildge, C. Nagel, C. Grun, Bronchoalveolar lavage in interstitial lung diseases: Does the recovery rate affect the results? *Respiration* **74**, 553–557 (2007).
 53. K. Koda, H. Hozumi, H. Yasui, Y. Suzuki, M. Karayama, K. Furuhashi, N. Enomoto, T. Fujisawa, N. Inui, Y. Nakamura, T. Suda, Predictors for bronchoalveolar lavage recovery failure in diffuse parenchymal lung disease. *Sci. Rep.* **11**, 1682 (2021).
 54. J. M. Löfdahl, K. Cederlund, L. Nathell, A. Eklund, C. M. Sköld, Bronchoalveolar lavage in COPD: Fluid recovery correlates with the degree of emphysema. *Eur. Respir. J.* **25**, 275–281 (2005).
 55. H. H. Uhlving, S. Mathiesen, F. Buchvald, K. Green, C. Heilmann, P. Gustafsson, K. Müller, K. G. Nielsen, Small airways dysfunction in long-term survivors of pediatric stem cell transplantation. *Pediatr. Pulmonol.* **50**, 704–712 (2015).
 56. S. Nyilas, L. Baumeler, M. Tamm, J. P. Halter, S. Savic, I. Korten, A. Meyer, F. Singer, J. R. Passweg, P. Latzin, D. Stolz, Inert gas washout in bronchiolitis obliterans following hematopoietic cell transplantation. *Chest* **154**, 157–168 (2018).
 57. G. G. King, J. Bates, K. I. Berger, P. Calverley, P. L. de Melo, R. L. Dellacà, R. Farré, G. L. Hall, I. Ioan, C. G. Irvin, D. W. Kaczka, D. A. Kaminsky, H. Kurosawa, E. Lombardi, G. N. Maksym, F. Marchal, B. W. Oppenheimer, S. J. Simpson, C. Thaminr, M. van den Berge, E. Oostveen, Technical standards for respiratory oscillometry. *Eur. Respir. J.* **55**, 1900753 (2020).
 58. M. S. De Assumpcao, E. da Silva Goncalves, M. S. Oliveira, J. D. Ribeiro, A. A. D. C. Toro, A. de Azevedo Barros-Filho, M. A. R. G. de Monteiro, C. I. S. Schivinski, Impulse oscillometry system and anthropometric variables of preschoolers, children and adolescents systematic review. *Curr. Pediatr. Rev.* **13**, 126–135 (2017).
 59. S. Stanojevic, B. L. Graham, B. G. Cooper, B. R. Thompson, K. W. Carter, R. W. Francis, G. L. Hall; Global Lung Function Initiative TLCO working group, Global Lung Function Initiative (GLI) TLCO, Official ERS technical standards: Global lung function initiative reference values for the carbon monoxide transfer factor for caucasians. *Eur. Respir. J.* **50**, 1700010 (2017).
 60. B. H. Culver, B. L. Graham, A. L. Coates, J. Wanger, C. E. Berry, P. K. Clarke, T. S. Hallstrand, J. L. Hankinson, D. A. Kaminsky, N. R. MacIntyre, M. C. McCormack, M. Rosenfeld, S. Stanojevic, D. J. Weiner; ATS Committee on Proficiency Standards for Pulmonary Function Laboratories, Recommendations for a standardized pulmonary function report. An official american thoracic society technical statement. *Am. J. Respir. Crit. Care Med.* **196**, 1463–1472 (2017).
 61. M. Koopman, P. Zanen, C. L. J. J. Kruitwagen, C. K. van der Ent, H. G. M. Arets, Reference values for paediatric pulmonary function testing: The Utrecht dataset. *Respir. Med.* **105**, 15–23 (2011).
 62. P. H. Quanjer, S. Stanojevic, T. J. Cole, X. Baur, G. L. Hall, B. H. Culver, P. L. Enright, J. L. Hankinson, M. S. M. Ip, J. Zheng, J. Stocks; ERS Global Lung Function Initiative, Multi-ethnic reference values for spirometry for the 3–95-yr age range: The global lung function 2012 equations. *Eur. Respir. J.* **40**, 1324–1343 (2012).
 63. A. B. Versluys, J. W. Rossen, B. van Ewijk, R. Schuurman, M. B. Bierings, J. J. Boelens, Strong association between respiratory viral infection early after hematopoietic stem cell transplantation and the development of life-threatening acute and chronic alloimmune lung syndromes. *Biol. Blood Marrow Transplant.* **16**, 782–791 (2010).
 64. M. S. Zinter, C. C. Dvorak, M. Y. Mayday, K. Iwanaga, N. P. Ly, M. E. McGarry, G. D. Church, L. E. Faricy, C. M. Rowan, J. R. Hume, M. E. Steiner, E. D. Crawford, C. Langelier, K. Kalantar, E. D. Chow, S. Miller, K. Shimano, A. Melton, G. A. Yanik, A. Sapru, J. L. DeRisi, Pulmonary metagenomic sequencing suggests missed infections in immunocompromised children. *Clin. Infect. Dis.* **68**, 1847–1855 (2019).
 65. M. Y. Mayday, L. M. Khan, E. D. Chow, M. S. Zinter, J. L. DeRisi, Miniaturization and optimization of 384-well compatible RNA sequencing library preparation. *PLOS ONE* **14**, e0206194 (2019).
 66. K. L. Kalantar, T. Carvalho, C. F. A. de Bourcy, B. Dimitrov, G. Dingle, R. Egger, J. Han, O. B. Holmes, Y.-F. Juan, R. King, A. Kislyuk, M. F. Lin, M. Mariano, T. Morse, L. V. Reynoso, D. R. Cruz, J. Sheu, J. Tang, J. Wang, M. A. Zhang, E. Zhong, V. Ahyong, S. Lay, S. Chea, J. A. Bohl, J. E. Manning, C. M. Tato, J. L. DeRisi, IDseq-An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *Gigascience* **9**, gaa111 (2020).
 67. S. Anders, D. J. McCarthy, Y. Chen, M. Okoniewski, G. K. Smyth, W. Huber, M. D. Robinson, Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protocol* **8**, 1765–1786 (2013).
 68. A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
 69. E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, A. Ma'ayan, Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128–128 (2013).
 70. K. J. Travaglini, A. N. Nabhan, L. Penland, R. Sinha, A. Gillich, R. V. Sit, S. Chang, S. D. Conley, Y. Mori, J. Seita, G. J. Berry, J. B. Shrager, R. J. Metzger, C. S. Kuo, N. Neff, I. L. Weissman, S. R. Quake, M. A. Krasnow, A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
 71. A. Panoskaltis-Mortari, M. Griese, D. K. Madtes, J. A. Belperio, I. Y. Haddad, R. J. Folz, K. R. Cooke; American Thoracic Society Committee on Idiopathic Pneumonia Syndrome, An official American Thoracic Society research statement: Noninfectious lung injury after hematopoietic stem cell transplantation: Idiopathic pneumonia syndrome. *Am. J. Respir. Crit. Care Med.* **183**, 1262–1279 (2011).
 72. M. H. Jagasia, H. T. Greinix, M. Arora, K. M. Williams, D. Wolff, E. W. Cowen, J. Palmer, D. Weisdorf, N. S. Treister, G. S. Cheng, H. Kerr, P. Stratton, R. F. Duarte, G. B. McDonald, Y. Inamoto, A. Vigorito, S. Arai, M. B. Dattiles, D. Jacobsohn, T. Heller, C. L. Kitko, S. A. Mitchell, P. J. Martin, H. Shulman, R. S. Wu, C. S. Cutler, G. B. Vogelsang, S. J. Lee, S. Z. Pavletic, M. E. Flowers, National institutes of health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. The 2014 diagnosis and staging working group report. *Biol. Blood Marrow Transplant.* **21**, 389–401 (2015).
 73. N. M. Davis, D. M. Proctor, S. P. Holmes, D. A. Relman, B. J. Callahan, Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6**, 226 (2018).
- Acknowledgments:** We thank D. Sheppard for the critical review of this manuscript. **Funding:** This study was funded by the National Marrow Donor Program Amy Strelzer Manasevit Grant (to M.S.Z.), NICHHD K12HD000850 (to M.S.Z.), NHLBI K23HL146936 (to M.S.Z.), an American Thoracic Society Foundation Grant (to M.S.Z.), NIAID F31AI150007 (to S.S.), NHLBI R01HL134828 (to M.A.M.), NHLBI R35HL140026 (to M.A.M.), and the Chan Zuckerberg Biohub (to J.L.D.). **Author contributions:** This study was conceptualized by M.S.Z., A.B.V., C.A.L., C.C.D., J.J.B., and J.L.D. Study methodology was derived from M.S.Z., A.B.V., C.A.L., M.S., E.K.F., S.P., C.C.D., J.J.B., and J.L.D. Study investigation was performed by M.S.Z., A.B.V., C.A.L., J.J.B., and J.L.D. Data visualization was conceived by M.S.Z., A.B.V., C.A.L., M.Y.M., S.S., M.S., J.J.B., and J.L.D. Manuscript draft writing was performed by M.S.Z., C.A.L., C.C.D., J.J.B., and J.L.D. Project administration, manuscript review, and manuscript editing were performed by M.S.Z., A.B.V., C.A.L., M.Y.M., G.R., S.S., M. Chan, E.K.F., M. Cancio, S.P., M.S., M.A.M., S.K., C.C.D., J.J.B., and J.L.D. **Competing interests:** C.A.L. participated in a data safety and monitoring board for ExCellThera and advisory boards for Bluebird Bio, Orchard Therapeutics, and Sobi. E.K.F. participated in an advisory board for Boehringer Ingelheim. M.A.M. participated in advisory boards for Johnson & Johnson and Citius Pharmaceuticals. C.C.D. participated in advisory boards for Jazz Pharmaceuticals, Alexion Inc., and Omeros Corp. J.J.B. participated in advisory boards for Race Oncology, Omeros, BlueRock, Advanced Clinical, Sanofi, and Avrobio. J.L.D. received consulting fees from Allen & Co. and PHC Inc. The other authors declare that they have no competing interests. **Data and materials availability:** All data associated with this study are present in the paper or the Supplementary Materials. Sequencing files are available through dbGaP study accession phs002208.v1.p1. Requests for data sharing should be addressed to matt.zinter@ucsf.edu.

Submitted 22 October 2021
 Accepted 28 January 2022
 Published 9 March 2022
 10.1126/scitranslmed.abm8646