Title

Chronic lung allograft dysfunction small airways reveal a lymphocytic inflammation gene signature

Journal

American Journal of Transplantation, 21(1)

ISSN

1600-6135

Authors

Dugger, Daniel T
Fung, Monica
Hays, Steven R
et al.

Publication Date

2021

DOI

10.1111/ajt.16293

Peer reviewed

Chronic Lung Allograft Dysfunction Small Airways Reveal A Lymphocytic Inflammation Gene Signature

Authors' Accepted Manuscript

American Journal of Transplantation

DOI: 10.1111/ajt.16293

Daniel T. Dugger[1,2], Monica Fung[1], Steven R. Hays[1], Jonathan P. Singer[1], Mary Ellen Kleinhenz[1], Lorriana E. Leard[1], Jeffrey A. Golden[1], Rupal J. Shah[1], Joyce S. Lee[3], Fred Deiter[1], Nancy Y. Greenland[2,4], Kirk D. Jones[4], Chaz R. Langelier[5], John R. Greenland[1,2]

[1] Department of Medicine, University of California, San Francisco, CA 94143
[2] Veterans Affairs Health Care System, San Francisco, CA 94121
[3] Department of Medicine, University of Colorado, Denver, CO 80045
[4] Department of Anatomic Pathology, University of California, San Francisco, CA 94143
[5] Chan Zuckerberg Biohub, San Francisco, CA 94143

Corresponding Author: John Greenland, john.greenland@ucsf.edu, 4150 Clement St, 111D, San Francisco, CA 94121; (415) 221-4810 ext. 22962.

ORCIDs:
0000-0002-2350-1743 (D. Dugger)
0000-0003-0224-7472 (J. Singer)
0000-0003-4454-6489 (F. Deiter)
0000-0002-1783-1711 (N. Greenland)
0000-0003-1422-8367 (J. Greenland)

Abbreviations:
B2M, Beta-2-Microglobulin
BAL, bronchoalveolar lavage
BOS, bronchiolitis obliterans syndrome
CF, cystic fibrosis
CI, confidence intervals
CLAD, Chronic lung allograft dysfunction
CT, computed tomography
CXCL9, C-X-C Motif Chemokine Ligand 9
DASH, Depletion of abundant sequences by hybridization
E-grade, lymphocytic bronchitis grade
EMT, epithelial to mesenchymal transition
FDR, false discovery rate
FEV_1, Forced expiratory volume in 1 second
FFPE, formalin-fixed paraffin embedded
FVC, Forced vital capacity
GO, Gene Ontology
HLA, human leukocyte antigen
ISHLT, International Society for Heart and Lung Transplantation
KEGG, Kyoto Encyclopedia of Genes and Genomes
LB, Lymphocytic bronchitis
LOESS, locally estimated scatterplot smoothing
MSigDB, Molecular Signatures Database
mTOR, mammalian target of rapamycin
PERMANOVA, permutational multivariate analysis of variance
RAS, restrictive allograft syndrome
RNAseq, next generation ribonucleic acid sequencing
ROC, receiver operating curve
rRNA, ribosomal ribonucleic acid
TBB, Transbronchial biopsy
TLC, Total Lung Capacity

Chronic Lung Allograft Dysfunction Small Airways Reveal A Lymphocytic Inflammation Gene Signature

ABSTRACT: Chronic lung allograft dysfunction (CLAD) is the major barrier to long-term survival following lung transplantation, and new mechanistic biomarkers are needed. Lymphocytic bronchitis (LB) precedes CLAD and has a defined molecular signature. We hypothesized that this LB molecular signature would be associated with CLAD in small airway brushings independent of infection. We quantified RNA expression from small airway brushings and transbronchial biopsies, using RNAseq and digital RNA counting, respectively, for 22 CLAD cases and 27 matched controls. LB metagene scores were compared across CLAD strata by Wilcoxon rank sum test. We performed unbiased host transcriptome pathway and microbial metagenome analysis in airway brushes and compared machine-learning classifiers between the two tissue types. This LB metagene score was increased in CLAD airway brushes ($P = 0.002$) and improved prediction of graft failure ($P = 0.02$). Gene expression classifiers based on airway brushes outperformed those using transbronchial biopsies. While infection was associated with decreased microbial alpha-diversity ($P \leq 0.04$), neither infection nor alpha-diversity was associated with LB gene expression. In summary, CLAD was associated with small airway gene expression changes not apparent in transbronchial biopsies in this cohort. Molecular analysis of airway brushings for diagnosing CLAD merits further examination in multicenter cohorts.

INTRODUCTION

Chronic lung allograft dysfunction (CLAD) limits quality and quantity of life following lung transplantation, affecting half of recipients as early as four years post-transplant (1, 2). Recently, two CLAD phenotypes have been recognized: bronchiolitis obliterans syndrome (BOS) and restrictive allograft syndrome (RAS) (3). BOS is focused within the small airways, while RAS also includes pleuroparenchymal fibrosis. Both are marked by irreversible decline in pulmonary function and histopathologic findings of extracellular matrix deposition. There are no therapies proven to prevent or reverse either subtype. Infections can also result in acute pulmonary function decline, and by the time CLAD diagnosis is confirmed, it may be too late to prevent irreversible fibrosis. Rapid CLAD identification could help to define clinical trial populations where CLAD was early enough to respond to treatment. Further, gene expression signatures associated with CLAD may inform potential therapies and could be used to confirm effective manipulation of specific molecular pathways.

Histopathologic examination of transbronchial biopsies is traditionally used to monitor for allograft rejection, but requires a large number of biopsies to diagnose rejection with a high degree of confidence (4). Small airway brushing has been proposed to diagnose inflammation indicative of rejection or infection (5) and can provide additional metagenomic data on the microbiome (6).

Lymphocytic inflammation in the large (bronchitis) and small (bronchiolitis) airways is associated with future development of CLAD (7-9). However, many centers do not collect large airway biopsies, and small airways are not always well represented on transbronchial biopsies. While moderate to severe lymphocytic inflammation on endobronchial biopsies is a rare finding, it is associated with substantial decreases in CLAD-free survival (7, 9). We recently described a gene expression signature based on RNA transcription changes in large airway brushings at the time of lymphocytic bronchitis (LB). This gene signature was validated in endobronchial and transbronchial biopsies in association with acute cellular rejection pathologies. However, the association between this LB gene signature

with CLAD pathology and its performance in small airway brushings are unknown (10). Here, we hypothesized that LB-associated gene expression would be increased in small airway brushings from subjects with CLAD.

METHODS

Cohort selection: We performed a case-control study nested within a longitudinal cohort of lung transplant recipients at the University of California, San Francisco who consented for small airway brushing. These subjects received immunosuppression and prophylactic therapies per institutional protocols as previously described (11).

We included all subjects with an airway brush within 3 months of CLAD onset. CLAD cases were identified by a ≥20% decline in FEV_1 from post-transplant baseline (2). Two investigators then reviewed each CLAD case and excluded cases with diagnostic uncertainty or alternative causes for FEV_1 decline (See Supplemental Figure 1). Cases were further classified as RAS, BOS, or mixed based on chart review of FVC, TLC, and CT imaging data using ISHLT criteria (3). Infection status was determined by presence of pathogenic microbes identified on BAL bacterial, fungal, and viral studies, understanding that some cases of asymptomatic colonization may have been classified as infection. Controls subjects were frequency matched at approximately 1:1 based on post-transplant time and BAL microbiology results. Further details on analysis, cohort matching, and immunosuppression are included in the Supplemental Methods.

Airway brushes and allograft biopsies: Lung transplant recipients who consented for airway brushing, allograft biopsy, and medical record review were sampled during standard-of-care bronchoscopies. Following bronchoalveolar lavage (BAL) and before biopsies, a cytology brush (Conmed #129) was advanced under fluoroscopic guidance into a basilar segment airway to about 3–4 cm from the periphery. The brush was agitated approximately 10 times and pulled back into the catheter. Brushes were stored in QIAzol lysis and preservation buffer (Qiagen #79306) on dry ice. After thawing and vortexing to dissociate epithelial cells from the brush, the lysate was passed through a QIAshredder (Qiagen #79656) and frozen at -80ºC prior to analysis. Transbronchial and endobronchial biopsies were performed as previously described (7). Two pathologists reassessed and regraded histopathologic features on coincident endobronchial and transbronchial biopsies in a blinded manner (7).

RNA sequencing: RNA was extracted from airway brushes using the Qiagen miRNeasy Mini Kit, and libraries generated using NEBNext Ultra II Library Prep Kit per manufacturer protocols (12) on an Agilent Bravo liquid handling instrument. Depletion of abundant sequences by hybridization (DASH) was employed to selectively deplete unwanted cDNA from human mitochondrial rRNA genes and enrich for host protein coding and microbial transcripts (13). RNAseq libraries underwent 150 nucleotide paired-end Illumina sequencing on a Novaseq 6000. Outliers were excluded based on principal component analysis using Tukey's fence criteria (k>3).

Digital RNA counting: RNA was extracted from formalin-fixed paraffin-embedded tissue blocks and quantified using the nanoString PanCancer Immune Profiling Panel, as previously described (10).

Analysis: Aligned RNAseq gene counts were normalized in DESeq and metagene values were calculated as the sum of gene counts normalized to a mean of 0 and standard deviation of 1. Differences in metagene score were compared by Wilcoxon rank sum test. Additional details on the analytic methods are included in the supplement.

RESULTS

Subject characteristics are shown in Table 1, with a subject enrollment flow diagram in Supplemental Figure 1. Across groups, recipients with CLAD were more commonly

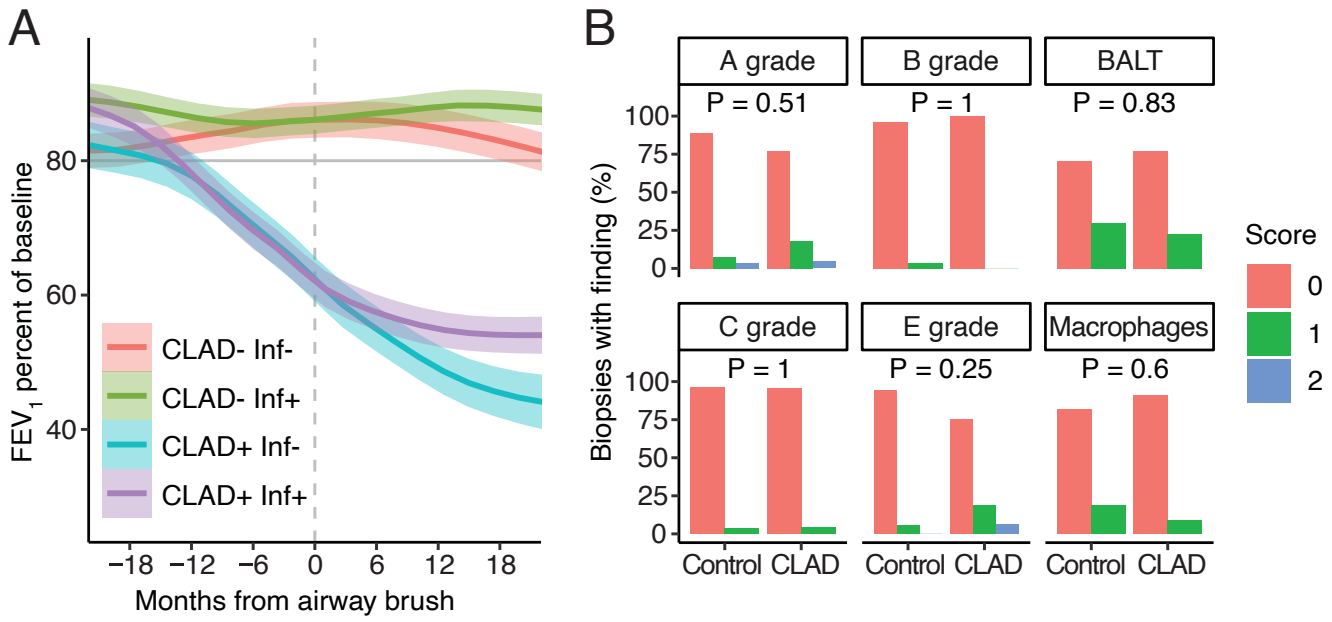


Figure 1: Histopathologic features fail to identify CLAD despite ongoing decline in pulmonary function. (A) FEV1 is shown as a smoothed function of time from airway brush for CLAD cases and controls with and without evidence of infection based on BAL bacterial and fungal cultures and viral PCR (Inf). (B) Histopathology review of transbronchial and endobronchial biopsies from subjects in both groups identified no distinguishing features between CLAD cases and controls. Grades refer to ISHLT criteria (30) with the addition of E-grade, as previously described for large airway inflammation. BALT, bronchial-associated lymphoid tissue. P-values are calculated by χ2-test.

female, and, as expected, overall survival and CLAD-free survival were worse in the CLAD- groups. Compared with controls, CLAD cases were more likely to undergo for cause bronchoscopy and had more cough and dyspnea, while subjects with infection were more likely to receive antimicrobials (Supplemental Table 1). Figure 1A shows the trajectory of FEV_1 in the CLAD and control groups. Airway brushes were collected at a median of 6 days (mean 111, interquartile range 1 – 40 days) after CLAD onset.

We did not observe any significant differences in histopathology on endobronchial or transbronchial biopsies associated with CLAD (Figure 1B) and this held true when controlling for infection status. Indeed, constrictive bronchiolitis (C-grade rejection) was evenly distributed between groups and there was only one case of ≥ mild lymphocytic bronchitis (E-grade rejection, analogous to B-grade but for large airway inflammation) (7).

LB-associated gene expression in CLAD. As our primary endpoint, we examined a previously-described LB metagene score in small airway brushing RNA, with a secondary comparison in transbronchial biopsy RNA (10). LB-associated gene expression was increased in CLAD subjects compared with controls by 0.87 standard deviations (95% CI 0.34 – 1.40, Figure 2A). However, there was no statistically significant difference when this LB gene expression score was calculated on transbronchial biopsies (delta 0.40, 95% CI -0.19 – 0.99 standard deviations). While infection could be expected to cause CLAD-independent airway inflammation, we observed no statistically significant differences when groups were stratified by infection status (P ≥0.29). Because RAS also involves the lung parenchyma, which is better represented in

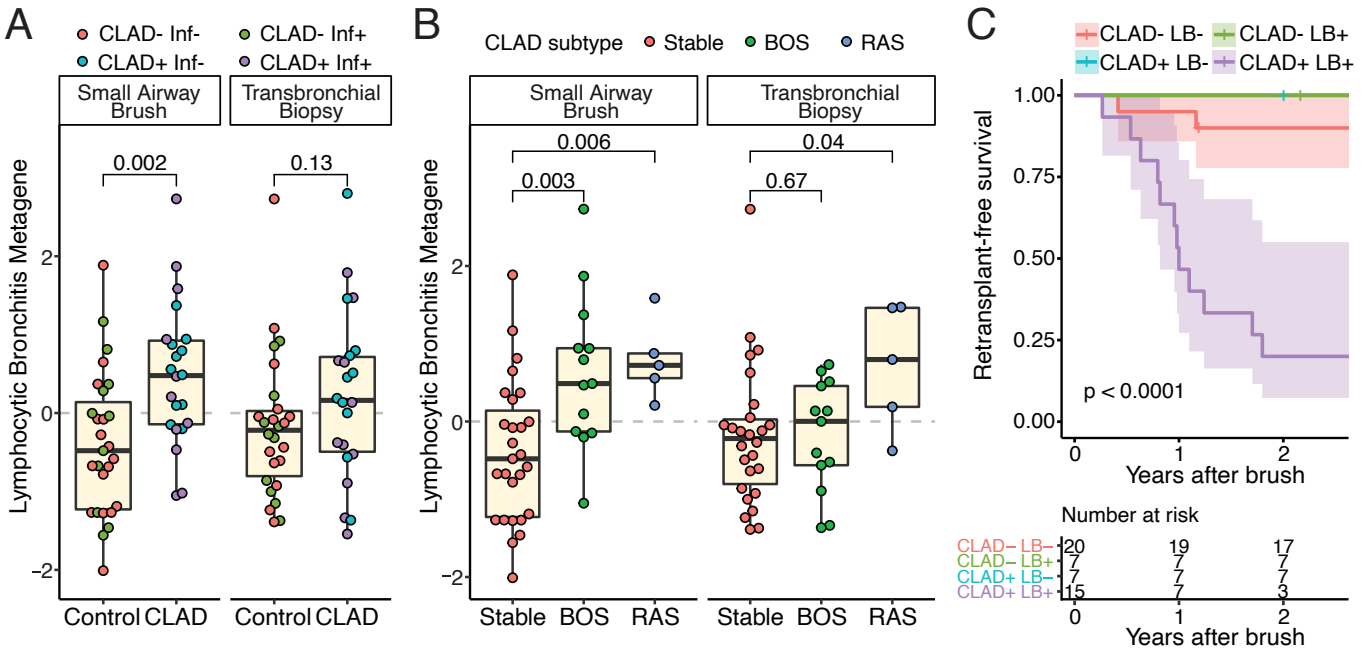


Figure 2: LB metagene expression is increased in CLAD and predicts allograft survival. Lymphocytic bronchitis metagene scores were calculated from RNA expression in small airway brushes and transbronchial biopsies and stratified by CLAD versus control (A). Infection is shown with green and purple points. No statistically significant differences were observed when stratified by infection (P = 0.57 and P = 0.29, respectively). (B) Metagene scores were grouped as Stable (N = 27), BOS (N = 13), or RAS (N = 5). Differences between groups were calculated by Wilcoxon rank sum test. (C) Kaplan–Meier plot showing time to graft failure minus date of airway brush stratified by CLAD status and LB metagene positivity, with the log-rank p-value shown.

transbronchial biopsies, we suspected that LB gene expression differences might be more apparent for transbronchial biopsies with RAS. Thus, in a secondary exploratory analysis, we looked at the subset of CLAD cases that were classified as BOS or RAS (excluding mixed CLAD). For small airway brushes, there were significant increases in LB-associated gene expression for both BOS and RAS (P ≤0.006). In transbronchial biopsies, expression was increased only for RAS (P = 0.04, Figure 2B).

LB expression and graft survival. We asked whether LB gene expression improved prediction of graft survival. Compared with a time to retransplant or death model including CLAD status and subject characteristics, adding the LB metagene resulted in a statically significantly improvement in fit (P = 0.02). Similarly, a standard deviation increase in LB metagene was associated with a 2.4-fold (95% CI 1.1 – 5.5) increased hazard of graft failure after adjustment for CLAD status and subject characteristics. Figure 2C shows graft survival time for subjects stratified by CLAD status and LB-metagene score >0.

Comparison of airway brushes to transbronchial biopsies. Because small airway brushes sample the site where constrictive bronchiolitis is focused, we hypothesized that gene expression-based assays on brushes would outperform those on transbronchial biopsies. Pathologist review of TBB revealed that airway was variably present, and minimally sufficient for the assessment of bronchiolitis in about one-third of cases. We thus compared differential expression between CLAD and non-CLAD samples for each gene assessed in both samples (Figure 3A). Interferon-related and other immune response genes (CXCL9, B2M, HLA-B), were upregulated in both groups (14). Overall, there was a positive correlation of CLAD-associated gene expression between airway brushes and transbronchial biopsies (P <0.001).

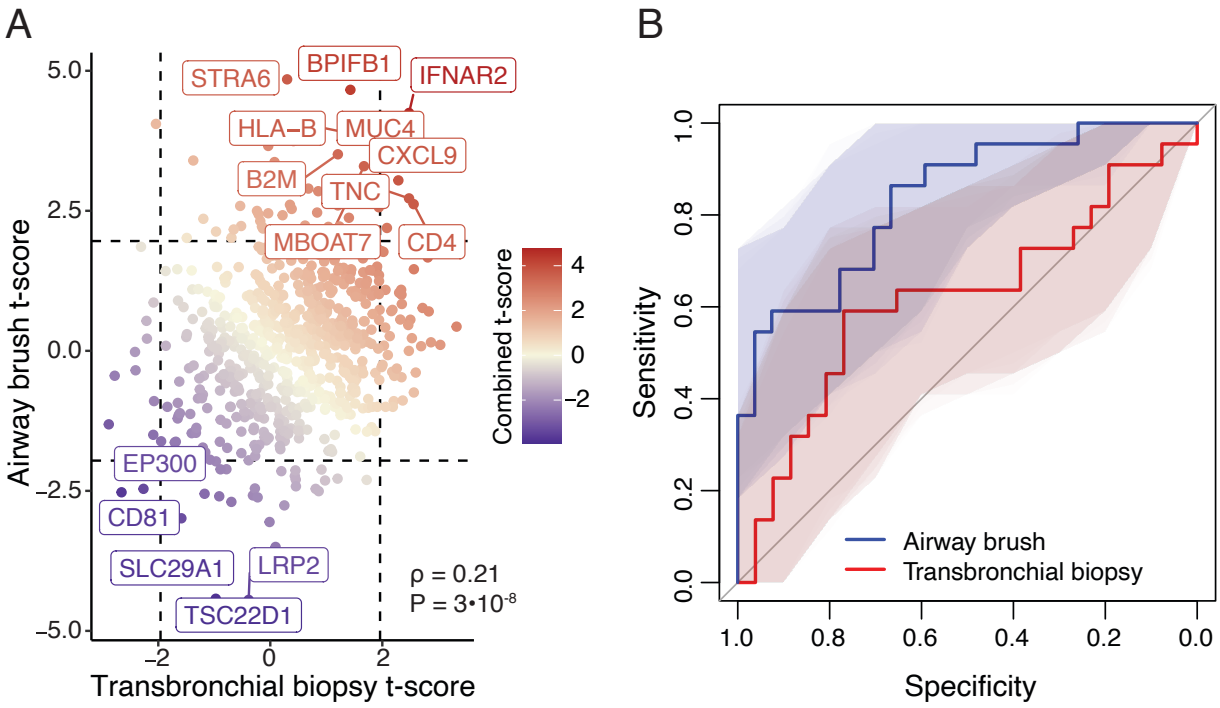


Figure 3: Comparison of CLAD-associated gene expression in transbronchial biopsy tissue and airway brushes. (A) Normalized gene expression values derived from airway brushes and biopsies were compared across CLAD strata by Student's t-test for each gene, with t-scores for airway brushes on the y-axis and transbronchial biopsies on the x-axis. Labeled genes were found to be significantly different in both contexts (Stouffer z-score >1.96). There was a statistically significant positive correlation between CLAD-versus control z-scores across tissue types (ρ=0.21, P = 3•10-8) (B) Receiver operating curve (ROC) analysis of predictions from leave-one-out cross validation using random forest models based on gene expression in airway brushes and transbronchial biopsies. The area under the ROC curve (AUC) was greater for airway brushes (AUC 0.84, 95% CI: 0.73–0.95) compared with the transbronchial biopsies (AUC 0.62, 95% CI 0.45–0.79, P = 0.04 by DeLong's ROC comparison test). Using an optimal cutoff of 0.57, airway brushes had sensitivity of 96%, specificity of 55%, positive predictive value of 72%, and negative predictive value of 92%. For transbronchial biopsies the optimal cut off was 0.51 with sensitivity 77%, specificity 59%, positive predictive value 70%, and negative predictive value 68%. The effect size CLAD versus control in brush random forest models was 1.00.

We used machine learning models to quantify the extent to which host gene expression in either tissue type could classify samples as CLAD or non-CLAD. Using a lasso-penalized logistic regression model, transbronchial biopsy expression data yielded an area under the receiver operating curve (AUC) of 0.49 (95% CI 0.44–0.55), while airway brush data yielded an AUC of 0.76 (95% CI 0.72–0.80; P-value <0.001 for difference in AUCs). Using random-forest models (Figure 3B), transbronchial biopsies had an AUC of 0.62 (95% CI 0.45–0.79) versus an AUC of 0.84 (95% CI 0.73–0.95) for airway brushes (P = 0.04 for AUC difference). A list of the top 50 genes ranked by importance in distinguishing CLAD versus non-CLAD is presented in Supplemental Table 2.

Transcriptional changes associated with CLAD. Next we performed unbiased differential gene expression analysis to determine other genes and pathways associated with CLAD in airway brushings from this cohort. We observed 38 genes upregulated and 26 genes downregulated with CLAD at a 10% FDR (Figure 4A). Hierarchical clustering analysis (Figure 4B) identified three gene expression groups: predominantly normal, a mixture of CLAD and infection, and a group with mostly CLAD samples. As shown in Figure 4C, CLAD was associated with a global shift in gene expression, whereas infection was not. Analysis of GO and KEGG pathways

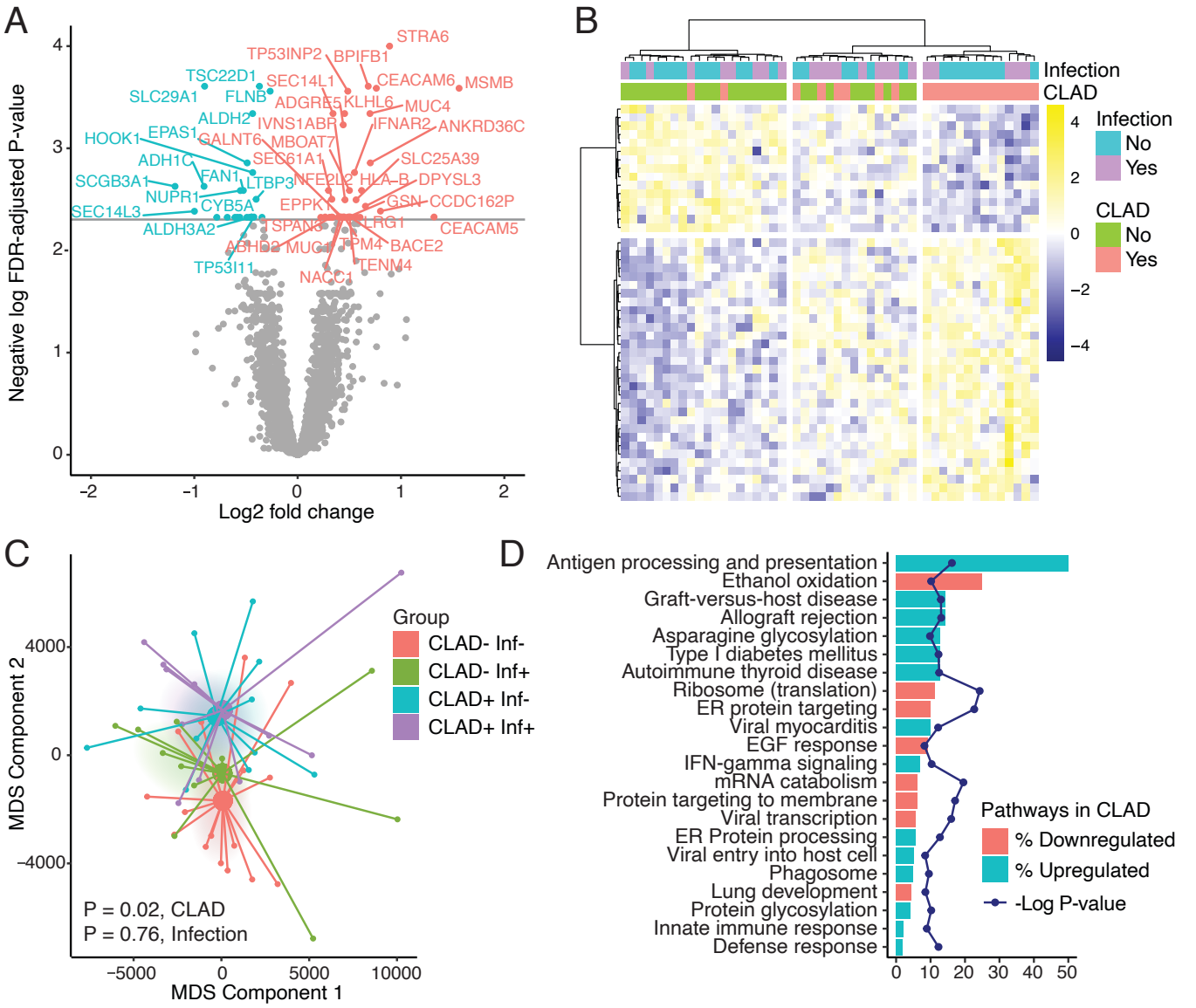


Figure 4: Transcriptome changes in CLAD. (A) Volcano plot labeled with the most differentially expressed genes between CLAD and CLAD-free samples. Labeled genes were upregulated (blue) or downregulated (red) in association with CLAD with a <0.1 FDR. (B) Heat map demonstrating gene expression for the 69 genes differentially expressed between CLAD and control samples at an FDR-adjusted P-value of <0.1. Unsupervised clustering identified 3 dominant clusters, with the right-most containing a preponderance of CLAD- samples. (C) Multidimensional scaling (MDS) plot of gene expression from airway brushes from 49 subjects stratified by CLAD and infection status. Separation between groups were calculated using PERMANOVA. (D) Gene ontology and KEGG pathway analyses of upregulated (blue) or downregulated (red) pathways in CLAD.

(Figure 4D), indicated that genes upregulated in CLAD were predominantly associated with immune pathways, including antigen presentation, allograft rejection, and interferon responses. Pathways downregulated during CLAD included protein synthesis, ethanol metabolism, EGF responses, and lung development genes.

To understand how molecular pathways differed with respect to time of CLAD onset, we examined smoothed MSigDB metagene scores (Supplemental Figure 2). Notch, Hedgehog, and Wnt/β-catenin pathways were greater in pre-CLAD samples, followed by hypoxia and angiogenesis. Prior to CLAD onset, there was an increase in mTORC1

signaling, while inflammatory pathways peaked coincident with CLAD onset. The LB metagene tracked with other late gene expression pathways, including interferon responses and cell cycle genes.

Microbiome. To understand impacts of the microbiome on LB-associated gene expression, we enumerated microbial transcripts in airway brushes to assess alpha- and beta-diversity, capitalizing on the RNA-seq reads in airway brushes that map to microbial genomes. Twenty-two subjects had a positive culture result. The most common microbes identified were *Aspergillus spp*., *Haemophilus parainfluenzae*, and *Pseudomonas aeruginosa* (Supplemental Table 3). We observed that within-individual (alpha) diversity metrics were decreased in samples with infection (Shannon, P = 0.01; Simpson, P = 0.04, Figure 5A), but there was no association between either metric of alpha-diversity and LB-associated gene expression (P ≥0.30). Additionally, there were no differences in alpha diversity when subjects were stratified by CLAD status (P ≥0.74).

We then assessed the types of species observed within groups (beta-diversity). Globally, we observed a trend towards statistically significant separation of groups when subjects were stratified by transplant indication (P = 0.08), but not by CLAD status (P = 0.15, Figure 5B). As shown in Supplemental Figure 2, there were two major clusters of microbial taxa, with an anaerobe-predominant cluster being linked to group D (fibrotic) lung transplant indications. When microbes were considered as broad categories (Figure 5C), we found a trend towards greater total microbe counts with group D indications, and particularly increased abundance of gram-negative facultative anaerobes. Groups C (cystic fibrosis) and D transplant indications had increased abundance of fungi and CF-associated pathogens (eg. *Pseudomonas*, *Pandoraea*, *Burkholderia*, etc.). Of note, *Pseudomonas* transcript abundance was increased in CF subjects without CLAD (log change 1.7, 95% CI 0.36 – 3.0), but not CF subjects with CLAD (log change -0.59, 95% CI -2.5 – 1.3). We examined differential abundance of bacterial genera across CLAD strata. The only genus associated with CLAD after FDR-adjustment was *Pseudomonas*, with a negative association (Figure 5D).

DISCUSSION

These data demonstrate a gene signature of LB to be increased in CLAD small airways versus controls and to identify those cases of CLAD at high risk for graft failure. Notably, histopathologic assessment of contemporaneous transbronchial biopsies showed no evidence of CLAD. While there were low numbers of subjects with RAS, transbronchial biopsies from these cases also demonstrated increased LB metagene scores. This provisional finding would be consistent with observations that RAS affects the parenchyma in addition to small airways, and thus would be better sampled by transbronchial biopsy (3). While there was a correlation between the CLAD-associated gene expression changes between the two tissue types, classifiers from gene expression data in small airway brushings outperformed classifiers based on transbronchial biopsy gene expression.

On the whole, CLAD was associated with upregulation of inflammatory gene pathways and recapitulated changes in secretory cells observed previously in chronic bronchitis, such as upregulation of *MSMB* and downregulation of *SCGB3A1* (15). Interferon activation has been linked to fibrosis and rejection (16, 17), and interferon-dependent genes such as *IFNAR2*, *CXCL9*, *HLA-B*, and *B2M*, were prominent in brushes and biopsies with CLAD (14). Indeed, the observed decreases in protein synthesis- and ethanol metabolism-associated genes may both be related to interferon signaling (18, 19). As shown in Supplemental Figure 2, gene expression pathways evolved over the course of CLAD, suggesting that future studies with pre-CLAD samples could identify a CLAD signature prior to CLAD onset. In particular, a loss of

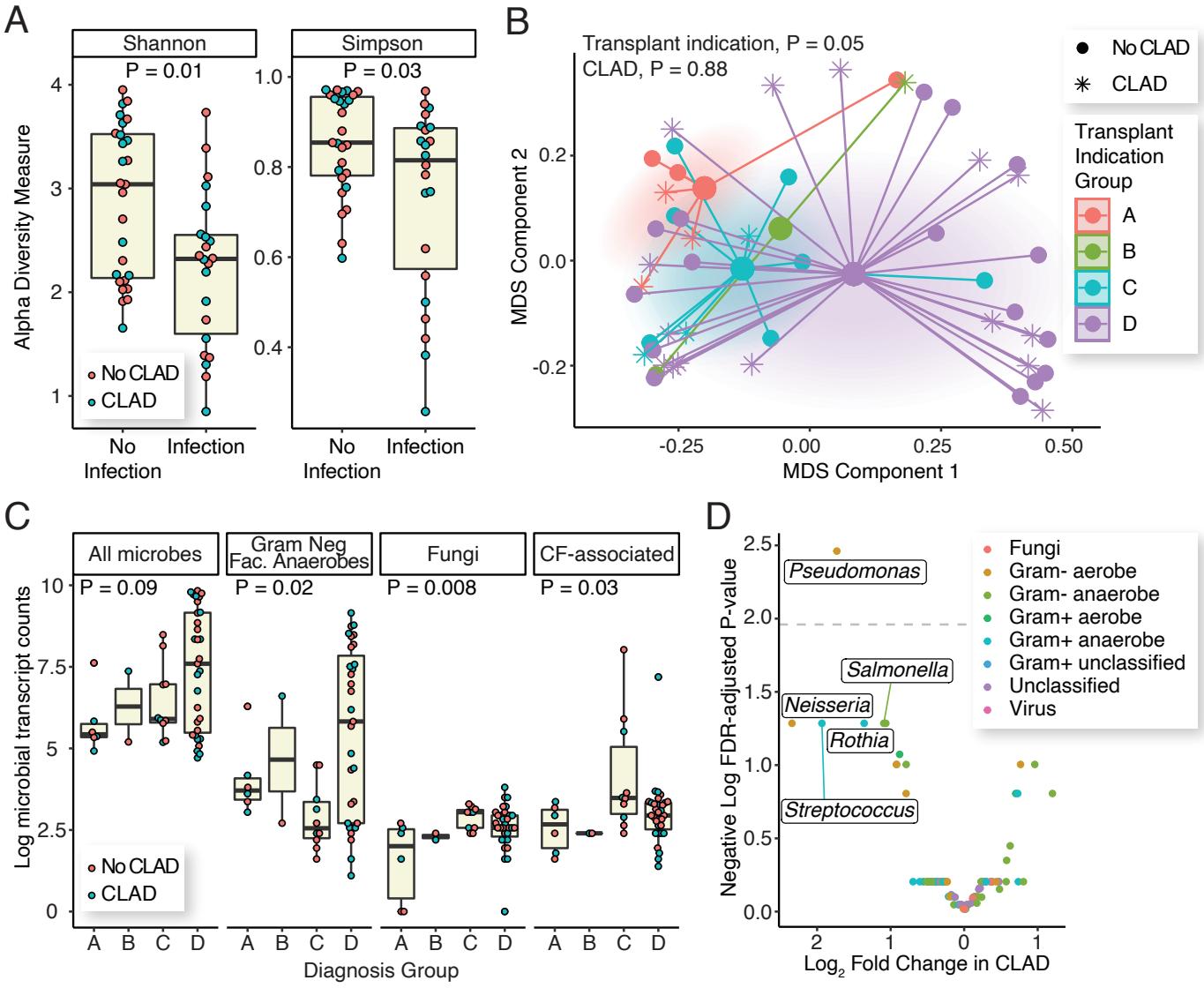


Figure 5: Metagenome analysis of microbiome in CLAD and non-CLAD airway brushes. (A) Alpha diversity was calculated using Shannon and Simpson metrics and stratified by infection status. There was decreased alpha-diversity associated with airway infection (P = 0.014 and P = 0.033, respectively), but no change in alpha-diversity associated with CLAD (P = 0.88 and P = 0.63), respectively. (B) Metagenomic beta-diversity was visualized using multidimensional scaling (MDS) of Bray–Curtis distances, with samples annotated by CLAD status and lung transplant indication groupings (A, Obstructive; B, Pulmonary vascular; C, Cystic fibrosis; and D, Fibrotic). Separation by group was determined by PERMANOVA, with P-values shown. (C) Sums of microbial transcript groups are shown log transformed and stratified by lung transplant indication. P-values were determined by ANOVA. There were no statistically significant differences by CLAD status across microbial transcript groups. (D) Differentially abundant genera were determined in CLAD versus non-CLAD samples and shown as negative-log of false-discovery rate (FDR)-adjusted P-value versus the log2 fold change associated with CLAD. The dashed line indicates FDR-adjusted alpha = 0.05 significance level.

homeostatic gene expression appears an early finding. There was downregulation of CD81 and LRP2 genes in both tissue types, which are both linked to airway homeostasis (20, 21). The increased effect size of the random forest model versus LB score suggests that CLAD might be better distinguished from normal using a score derived from CLAD cases and controls.

Although infection status did not affect host transcriptome, it was the major determinant of metagenomic alpha-diversity, consistent with prior studies (12, 22). Importantly, LB metagene scores were independent of clinical infection status and alpha-diversity metrics. The absence of significant viral transcription argues against occult viral infection as the driver of this gene score, despite the predominance of interferon-associated transcripts.

The increased abundance of *Pseudomonas* in CF and CLAD-free subjects is consistent with prior reports that reestablishment of pre-transplant flora is associated with protection from CLAD in individuals with CF. Mechanistically, this protection might result from suppressed airway inflammatory responses or microbial strain differences (23). The observed increased incidence of gram-negative facultative anaerobes in subjects with pulmonary fibrosis may indicate aspiration of oral flora. Aspiration of gastric fluid may be a risk factor for CLAD, and increased aspiration has been observed in lung transplant recipients with pulmonary fibrosis (24).

While the present data demonstrate a LB gene signature in the context of CLAD, there are several important limitations. Both biopsy and brushing are subject to sampling error. In particular, obliterated airways may be inaccessible to cytology brushes. Brushing may be less susceptible to sampling error if there is a “field of injury” beyond that affected airways, as has been reported with lung cancer (25). Findings might also be dependent on the proportion of surveillance versus for-cause bronchoscopies, although the inclusion of both reflects clinical practice. It is not known how these gene expression patterns would differ at other centers, where subject characteristics and immunosuppression strategies might differ. As this study targeted early CLAD cases, kinetic data included a paucity of late-CLAD samples. We observed downregulation of EGFR in CLAD in contradistinction with recently published findings on the Amphiregulin pathway in CLAD (26). However, this downregulation was limited to subjects with early CLAD, and this difference could reflect the enrichment of early CLAD cases in our cohort. While digital RNA counting is relatively robust when assessing FFPE RNA, it is not known to what extent differences between the two techniques contributed to the observed differences in classifier accuracy. Although most cells collected by airway brushing are epithelial cells, leukocytes are also present (27). We did not perform differential analysis to define the populations of cells gathered by brushings but recognize that small numbers of infiltrating leukocytes could result in highly differentially expressed gene counts. Future experiments using single cell sequencing could help identify the cell types responsible for the observed gene expression changes in CLAD. We did not assess if protein concentration data corroborated the observed transcriptional differences. We have found that gene expression and protein translation are only correlated for some genes in airway epithelial cells (23). While this transcriptional signature may be a useful biomarker for CLAD, investigations into potential therapeutic targets would need to start with assessment of protein correlates across tissue compartments. Finally, Illumina-based RNAseq technology may not be optimal for reduction to practice, since this technology becomes impractical without pooling of multiple samples. While pooling could be accomplished through a central lab, substitution of other technologies, such as quantitative PCR or Nanopore sequencing might be advantageous for rapid diagnostics (28).

In summary, gene expression analysis of small airway brushings has the potential to facilitate the diagnosis of CLAD, while simultaneously assessing for airway infection. If substituted for transbronchial biopsies, airway brushing could also reduce risk to patients (29). While transbronchial biopsies are necessary for establishing clinically actionable diagnoses of acute cellular rejection, our prior data showed similar gene expression changes associated with acute cellular rejection (10). Airway

brushing could also identify key CLAD pathobiologies leading to targeted therapies. Infection can be identified by alpha-diversity, even with relatively low coverage of the microbial transcriptome. Early signatures could also identify subjects at increased CLAD risk for clinical trials of preventive therapy and could be used as a surrogate measure to shorten the timeframe of such studies. While much work is needed before such diagnostics could be implemented clinically, this study demonstrates ways in which airway brushing analyses could improve management for lung transplant recipients.

Acknowledgments:

Author Contributions: D.T.D., M.F., C.R.L., and J.R.G. designed the experiments. D.T.D. and M.F. performed experiments. S.R.H., J.P.S., L.E.L., J.A.G., R.J.S., J.S.L. and M.E.K. obtained and provided clinical samples for RNA sequencing, metagenome, and culture data. D.T.D., M.F., F.D., C.R.L., and J.R.G. performed data analysis. J.R.G. wrote bioinformatics R code. N.Y.G. and K.D.J. provided histopathology grading and interpretation. D.T.D. and J.R.G. wrote the manuscript. All authors participated in manuscript revisions.

Funding: Clinical Sciences Research & Development Service of the Veterans Affairs Office of Research and Development (CX001034 & CX002011, J.R.G.), the UCSF Department of Anatomic Pathology (N.Y.G.), the Nina Ireland Program for Lung Health, the Cystic Fibrosis Foundation Therapeutics (GREENL16XX0), and the National Institutes of Health R01HL151552 (J.R.G) and K23HL138461 (C.R.L.).

TABLE 1: Subject Characteristics

	CLAD- Inf-	CLAD- Inf+	CLAD+ Inf-	CLAD+ Inf+	P-value
N	16	11	11	11	
Recipient age, mean (SD)	52.1 (20.6)	49.1 (16.4)	53.5 (9.2)	47.5 (12.1)	0.80
Donor age, mean (SD)	36.2 (14.8)	33.5 (14.5)	29.5 (12.9)	36.0 (14.0)	0.64
Male recipient, N (%)	10 (62.5)	10 (90.9)	3 (27.3)	7 (63.6)	0.02
Male donor, N (%)	11 (68.8)	8 (72.7)	9 (81.8)	8 (72.7)	0.90
Recipient ethnicity, N (%)					0.11
White	12 (75.0)	8 (72.7)	5 (45.5)	7 (63.6)	
Black	0 (0)	2 (18.2)	0 (0.0)	0 (0)	
Hispanic	3 (18.8)	1 (9.1)	4 (36.4)	4 (36.4)	
Other	1 (6.2)	0 (0)	2 (18.2)	0 (0)	
Donor ethnicity, N (%)					0.30
White	6 (37.5)	7 (63.6)	8 (72.7)	6 (54.5)	
Black	4 (25.0)	0 (0.0)	1 (9.1)	4 (36.4)	
Hispanic	3 (18.8)	3 (27.3)	1 (9.1)	0 (0)	
Other	3 (18.8)	1 (9.1)	1 (9.1)	1 (9.1)	
Diagnosis group, N (%)					0.89
A - Obstructive	2 (12.5)	1 (9.1)	2 (18.2)	1 (9.1)	
B - Pulmonary Vascular	0 (0.0)	1 (9.1)	0 (0.0)	1 (9.1)	
C - CF	4 (25.0)	3 (27.3)	1 (9.1)	2 (18.2)	
D - Restrictive	10 (62.5)	6 (54.5)	8 (72.7)	7 (63.6)	
Brush post-transplant years, mean (SD)	3.54 (3.97)	2.83 (3.56)	3.48 (3.41)	3.76 (2.62)	0.93
Lung allocation score, mean (SD)	63.5 (23.5)	57.1 (18.9)	62.2 (24.8)	53.1 (22.6)	0.67
Double lung transplant, N (%)	12 (80.0)	10 (90.9)	11 (100)	10 (90.9)	0.42
Gastric fundoplication, N(%)					0.78
Prior to brush	1 (6.2)	2 (18.2)	2 (18.2)	1 (9.1)	
Post-brush	2 (12.5)	1 (9.1)	0 (0)	2 (18.2)	
Mycophenolic acid, mg per day (mean, SD)	750 (665)	1040 (753)	363 (377)	885 (734)	0.11
CLAD type					0.53
Obstructive (BOS)			7 (63.6)	6 (54.5)	
Mixed			1 (9.1)	3 (27.3)	
Restrictive (RAS)			3 (27.3)	2 (18.2)	
Re-transplant after brush (%)	1 (6.2)	0 (0)	1 (9.1)	2 (18.2)	0.47
CLAD-free survival years, restricted mean (se)	10.2 (1.4)	8.5 (1.6)	3.2 (0.9)	3.5 (0.8)	<0.001
Overall survival years, mean (se)	13.8 (1.1)	15.5	7.1 (1.8)	11.5 (1.9)	0.003

REFERENCES:

1. Kulkarni HS, Cherikh WS, Chambers DC, Garcia VC, Hachem RR, Kreisel D et al. Bronchiolitis obliterans syndrome-free survival after lung transplantation: An International Society for Heart and Lung Transplantation Thoracic Transplant Registry analysis. J Heart Lung Transplant 2019;38(1):5-16.

2. Verleden GM, Glanville AR, Lease ED, Fisher AJ, Calabrese F, Corris PA et al. Chronic lung allograft dysfunction: Definition, diagnostic criteria, and approaches to treatment-A consensus report from the Pulmonary Council of the ISHLT. J Heart Lung Transplant 2019;38(5):493-503.

3. Glanville AR, Verleden GM, Todd JL, Benden C, Calabrese F, Gottlieb J et al. Chronic lung allograft dysfunction: Definition and update of restrictive allograft syndrome-A consensus report from the Pulmonary Council of the ISHLT. J Heart Lung Transplant 2019;38(5):483-492.

4. Scott JP, Fradet G, Smyth RL, Mullins P, Pratt A, Clelland CA et al. Prospective study of transbronchial biopsies in the management of heart-lung and single lung transplant patients. J Heart Lung Transplant 1991;10(5 Pt 1):626-636; discussion 636-627.

5. Chambers DC, Hodge S, Hodge G, Yerkovich ST, Kermeen FD, Reynolds P et al. A novel approach to the assessment of lymphocytic bronchiolitis after lung transplantation--transbronchial brush. J Heart Lung Transplant 2011;30(5):544-551.

6. Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D et al. Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. Int J Cancer 2018;142(4):769-778.

7. Greenland JR, Jones KD, Hays SR, Golden JA, Urisman A, Jewell NP et al. Association of large-airway lymphocytic bronchitis with bronchiolitis obliterans syndrome. Am J Respir Crit Care Med 2013;187(4):417-423.

8. Glanville AR, Aboyoun CL, Havryk A, Plit M, Rainer S, Malouf MA. Severity of lymphocytic bronchiolitis predicts long-term outcome after lung transplantation. Am J Respir Crit Care Med 2008;177(9):1033-1040.

9. Verleden SE, Scheers H, Nawrot TS, Vos R, Fierens F, Geenens R et al. Lymphocytic bronchiolitis after lung transplantation is associated with daily changes in air pollution. Am J Transplant 2012;12(7):1831-1838.

10. Greenland JR, Wang P, Brotman JJ, Ahuja R, Chong TA, Kleinhenz ME et al. Gene signatures common to allograft rejection are associated with lymphocytic bronchitis. Clin Transplant 2019;33(5):e13515.

11. Greenland JR, Chong T, Wang AS, Martinez E, Shrestha P, Kukreja J et al. Suppressed calcineurin-dependent gene expression identifies lung allograft recipients at increased risk of infection. Am J Transplant 2018;18(8):2043-2049.

12. Langelier C, Kalantar KL, Moazed F, Wilson MR, Crawford ED, Deiss T et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. Proc Natl Acad Sci U S A 2018;115(52):E12353-E12362.

13. Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H et al. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol 2016;17:41.

14. Tan H, Derrick J, Hong J, Sanda C, Grosse WM, Edenberg HJ et al. Global transcriptional profiling demonstrates the combination of type I and type II interferon enhances antiviral and immune responses at clinically relevant doses. J Interferon Cytokine Res 2005;25(10):632-649.

15. Wang G, Lou HH, Salit J, Leopold PL, Driscoll S, Schymeinsky J et al. Characterization of an immortalized human small airway basal stem/progenitor cell line with airway region-specific differentiation capacity. Respir Res 2019;20(1):196.

16. Rascio F, Pontrelli P, Accetturo M, Oranger A, Gigante M, Castellano G et al. A type I interferon signature characterizes chronic antibody-mediated rejection in kidney transplantation. J Pathol 2015;237(1):72-84.

17. Wu M, Skaug B, Bi X, Mills T, Salazar G, Zhou X et al. Interferon regulatory factor 7 (IRF7) represents a link between inflammation and fibrosis in the pathogenesis of systemic sclerosis. Ann Rheum Dis 2019;78(11):1583-1591.

18. Osna NA, Clemens DL, Donohue TM, Jr. Ethanol metabolism alters interferon gamma signaling in recombinant HepG2 cells. Hepatology 2005;42(5):1109-1117.

19. Gupta R, Kim S, Taylor MW. Suppression of ribosomal protein synthesis and protein translation factors by Peg-interferon alpha/ribavirin in HCV patients blood mononuclear cells (PBMC). J Transl Med 2012;10:54.

20. Jin Y, Takeda Y, Kondo Y, Tripathi LP, Kang S, Takeshita H et al. Double deletion of tetraspanins CD9 and CD81 in mice leads to a syndrome resembling accelerated aging. Sci Rep 2018;8(1):5145.

21. Vohwinkel CU, Buchackert Y, Al-Tamari HM, Mazzocchi LC, Eltzschig HK, Mayer K et al. Restoration of Megalin-Mediated Clearance of Alveolar Protein as a Novel Therapeutic Approach for Acute Lung Injury. Am J Respir Cell Mol Biol 2017;57(5):589-602.

22. Langelier C, Zinter MS, Kalantar K, Yanik GA, Christenson S, O'Donovan B et al. Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. Am J Respir Crit Care Med 2018;197(4):524-528.

23. Dugger DT, Fung M, Zlock L, Caldera S, Sharp L, Hays SR et al. Cystic Fibrosis Lung Transplant Recipients Have Suppressed Airway Interferon Responses During *Pseudomonas* Infection. Cell Reports Medicine 2020;1(4):100055.

24. Patti MG, Vela MF, Odell DD, Richter JE, Fisichella PM, Vaezi MF. The Intersection of GERD, Aspiration, and Lung Transplantation. J Laparoendosc Adv Surg Tech A 2016;26(7):501-505.

25. Silvestri GA, Vachani A, Whitney D, Elashoff M, Porta Smith K, Ferguson JS et al. A Bronchial Genomic Classifier for the Diagnostic Evaluation of Lung Cancer. N Engl J Med 2015;373(3):243-251.

26. Todd JL, Kelly FL, Nagler A, Banner K, Pavlisko EN, Belperio JA et al. Amphiregulin contributes to airway remodeling in chronic allograft dysfunction after lung transplantation. Am J Transplant 2019.

27. Forrest IA, Murphy DM, Ward C, Jones D, Johnson GE, Archer L et al. Primary airway epithelial cell culture from lung transplant recipients. Eur Respir J 2005;26(6):1080-1085.

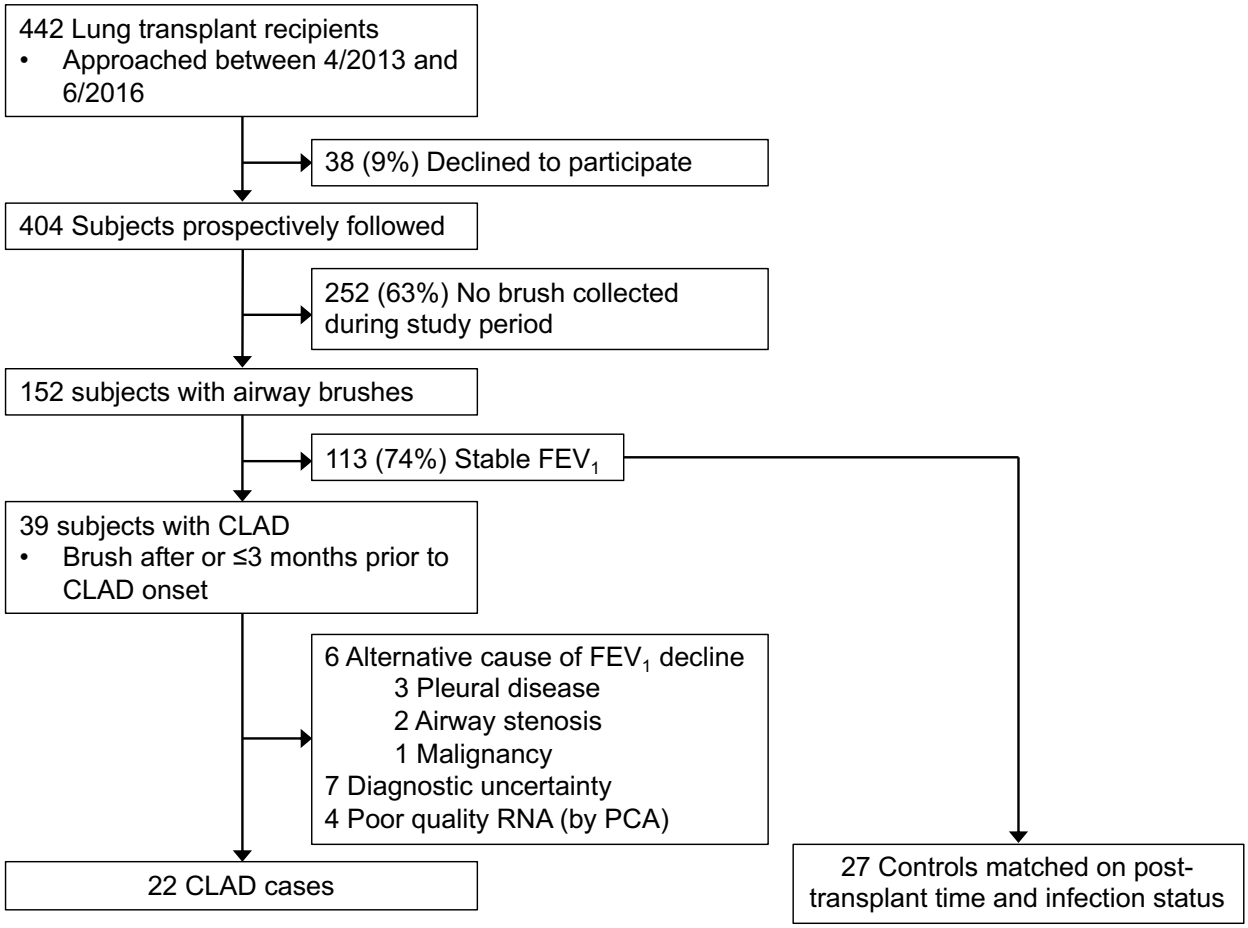
28. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med 2015;7:99.

29. Rademacher J, Suhling H, Greer M, Haverich A, Welte T, Warnecke G et al. Safety and efficacy of outpatient bronchoscopy in lung transplant recipients - a single centre analysis of 3,197 procedures. Transplant Res 2014;3:11.

Chronic Lung Allograft Dysfunction Small Airways Reveal A Lymphocytic Inflammation Gene Signature

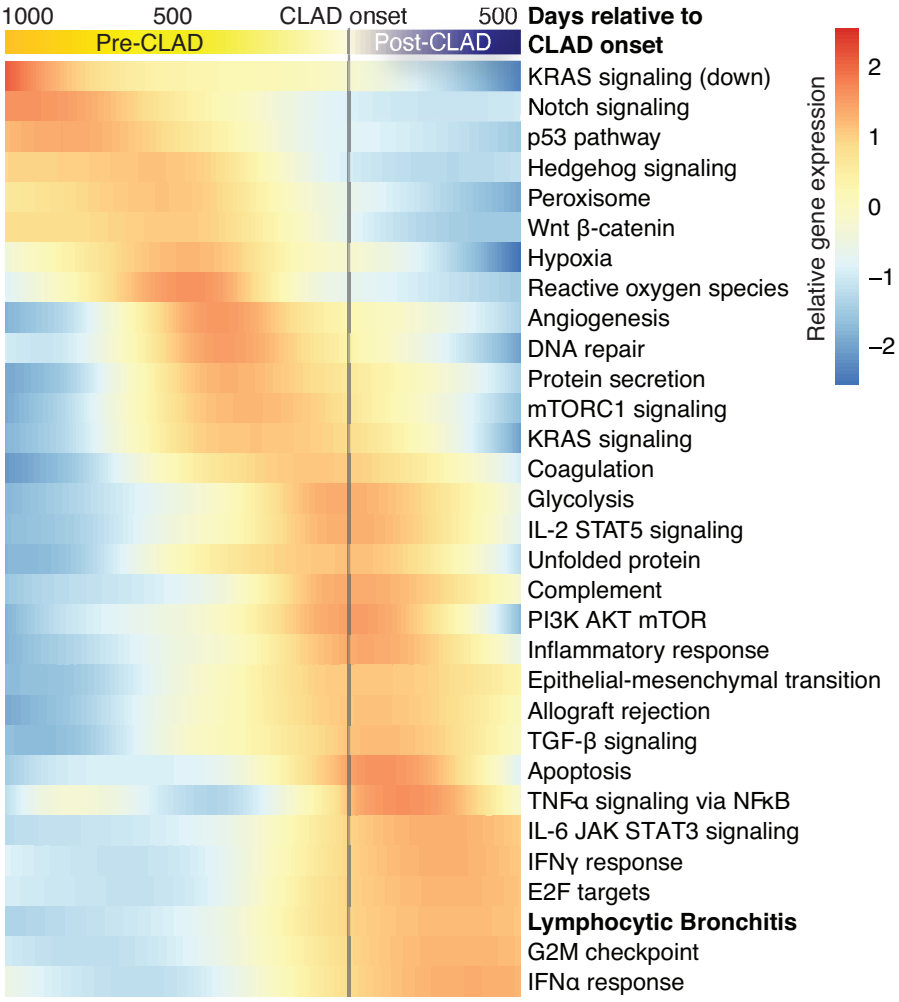
Supplemental Figures

Supplemental Figure 1: Study enrollment flow diagram



Airway brushings were collected on consenting subjects in a longitudinal cohort study at UCSF. During the period, 404 of the 442 subjects approached consented to an airway brush collected for research during clinically indicated bronchoscopies. Within this group of consenting lung transplant recipients, 152 underwent a procedure (for-cause or surveillance) to obtain BAL, airway, and/or biopsy samples and had an airway brush collected. At the time of airway brushing, 117 of the subjects had stable pulmonary function testing while 39 had a ≥20% decline in FEV_1 from post-transplant baseline, consistent with a diagnosis of CLAD. Upon medical record review of these 39 cases, 13 subjects were excluded from the CLAD group because of identification of an alternative causes of declining lung function or diagnostic uncertainty. Four CLAD subjects were excluded due to insufficient RNA quality to perform RNA-sequencing. From the cohort of subject who received airway brushing during this study period, we then selected control cases in approximately 1:1 ratio based upon time post-transplant and infection status. Subject demographics are presented in Table 1.

Supplemental Figure 2: Pathway expression over time



Relative expression of MSigDB Hallmark pathway metagenes and the lymphocytic bronchitis metagene are shown as a LOESS-function of airway brush time relative to CLAD onset (x-axis). Pathways at the top had relatively earlier expression (or later suppression), while pathways at the bottom of the list had peak expression after CLAD. In individuals not observed to develop CLAD or long before CLAD onset, airway brushes show increased expression of the Hedgehog pathway. It has been reported that hedgehog signaling is required to maintain quiescence within the airway epithelium (1). Similarly, Notch signaling is reported as critical for maintaining airway epithelial integrity and is lost early in the pre-CLAD period (2). Although Wnt/β-catenin signaling pathways have been linked to airway homeostasis and fibrosis, in this context, this pathway also appears to be associated with airway homeostasis, possibly reflecting airway regeneration in the context of chronic alloimmune injury (3). Hypoxia and angiogenesis pathways peak around 1–2 years prior to CLAD onset, potentially reflecting microvascular loss in the context of abnormal bronchial artery circulation (4). DNA damage pathways peaked next, which could relate to critical telomere shortening that is a risk factor for CLAD (5). Interestingly, mTORC1 signaling was observed immediately prior to CLAD onset, and mTORC1 inhibition has been investigated as a therapy to prevent CLAD (6). CLAD onset was associated with cell cycle changes, including G2M and E2F pathways, as well as epithelial to mesenchymal transition, which has also been reported in CLAD (6). These dynamic gene expression pathway changes may suggest that the gene signatures most useful for identifying patients prior to CLAD onset may differ from those optimal for diagnosing CLAD.

Supplemental Table 1: Indications for Bronchoscopy and Interventions

	CLAD- Inf-	CLAD- Inf+	CLAD+ Inf-	CLAD+ Inf+	P-value
Number of subjects	16	11	11	11	
Indication					
Surveillance	8 (50.0)	6 (54.5)	1 (9.1)	4 (36.4)	0.106
Acute symptoms	6 (37.5)	2 (18.2)	6 (54.5)	4 (36.4)	0.371
Follow up rejection infection	4 (25.0)	5 (45.5)	5 (45.5)	4 (36.4)	0.644
Airway issue stent	0 (0.0)	0 (0.0)	0 (0.0)	1 (9.1)	0.317
Dyspnea	3 (18.8)	1 (9.1)	6 (54.5)	5 (45.5)	0.056
Cough	4 (25.0)	0 (0.0)	5 (45.5)	5 (45.5)	0.056
Flu like symptoms	0 (0.0)	1 (9.1)	1 (9.1)	0 (0.0)	0.465
Fatigue	2 (12.5)	0 (0.0)	3 (27.3)	1 (9.1)	0.267
Decreased peak flow or FEV_1	5 (31.2)	5 (45.5)	6 (54.5)	6 (54.5)	0.565
Treatment					
Solumedrol (500mg IV x3days)	0 (0.0)	0 (0.0)	2 (18.2)	0 (0.0)	0.066
Prednisone taper	0 (0.0)	1 (9.1)	3 (27.3)	1 (9.1)	0.148
Increased chronic immunosuppression	0 (0.0)	1 (9.1)	0 (0.0)	1 (9.1)	0.465
Treated for Infection	0 (0.0)	8 (72.7)	0 (0.0)	10 (90.9)	<0.001

Supplemental Table 2: Top 50 genes sorted by importance in random forest models of CLAD versus stable

Gene Name	Increase in node purity (median IQR 25%–50%)				Mean standard error % (median IQR 25%–50%)			
	All genes		nanoString restricted		All genes		nanoString restricted	
STRA6	0.40	0.35 – 0.47	1.22	1.15 – 1.38	4.20	3.80 – 4.60	8.88	8.49 – 9.38
CDC42EP5	0.29	0.23 – 0.34			3.01	2.24 – 3.31		
LRP2	0.08	0.06 – 0.12	0.38	0.33 – 0.46	1.25	0.11 – 1.74	3.10	2.58 – 3.64
IFNAR2	0.08	0.05 – 0.13	0.41	0.34 – 0.48	1.02	0.42 – 1.59	3.94	3.40 – 4.45
CEACAM6	0.09	0.06 – 0.13	0.39	0.32 – 0.47	1.79	1.38 – 2.02	3.94	3.12 – 4.21
TP53INP2	0.07	0.05 – 0.09	0.39	0.33 – 0.45	1.33	0.94 – 1.62	3.75	3.07 – 4.33
BPIFB1	0.07	0.03 – 0.09	0.34	0.29 – 0.43	1.02	0.43 – 1.41	3.44	2.91 – 4.10
GSN	0.05	0.03 – 0.07	0.33	0.29 – 0.41	0.62	-0.14 – 1.22	3.06	2.67 – 3.54
TSC22D1	0.05	0.03 – 0.09	0.32	0.27 – 0.38	1.00	0.13 – 1.53	3.41	2.82 – 3.87
TFEB	0.04	0.03 – 0.05	0.32	0.26 – 0.38	1.00	0.35 – 1.42	3.64	3.24 – 4.23
UHRF1	0.04	0.02 – 0.06	0.27	0.23 – 0.32	1.00	-0.05 – 1.22	2.52	1.91 – 3.12
ALDH2	0.16	0.12 – 0.19			2.10	1.68 – 2.68		
SLC29A1	0.02	0.01 – 0.03	0.22	0.19 – 0.26	1.00	0.00 – 1.00	2.95	2.33 – 3.46
NOL3	0.11	0.09 – 0.15			1.74	1.10 – 2.20		
NFE2L2	0.02	0.01 – 0.03	0.20	0.15 – 0.24	1.00	0.00 – 1.32	2.71	2.26 – 3.15
CXCL9	0.02	0.01 – 0.03	0.15	0.13 – 0.19	0.00	-0.63 – 1.00	1.03	0.46 – 1.71
DHX16	0.02	0.02 – 0.05	0.18	0.13 – 0.23	1.00	-0.81 – 1.00	1.73	1.17 – 2.34
ST6GAL1	0.02	0.01 – 0.03	0.16	0.13 – 0.20	1.00	0.00 – 1.00	2.89	2.30 – 3.19
WWC3	0.10	0.07 – 0.15			1.45	0.88 – 1.84		
CSNK1A1	0.09	0.06 – 0.12			1.35	0.76 – 1.83		
HIST1H4B	0.08	0.06 – 0.12			1.59	1.18 – 1.88		
CCDC127	0.07	0.06 – 0.10			1.33	0.61 – 1.94		
TPM4	0.02	0.01 – 0.03	0.15	0.11 – 0.18	0.00	-1.00 – 1.00	1.33	0.78 – 1.85
NUPR1	0.09	0.06 – 0.12			1.59	1.15 – 2.28		
DYNC1LI1	0.07	0.06 – 0.08			0.90	0.10 – 1.41		
AEN	0.08	0.05 – 0.11			1.24	0.65 – 1.79		
MSMB	0.09	0.05 – 0.12			1.56	1.00 – 2.06		
NDRG1	0.07	0.05 – 0.10			1.32	0.85 – 1.66		
RBM5-AS1	0.06	0.04 – 0.08			0.99	-0.01 – 1.34		
NFATC3	0.00	0.00 – 0.01	0.10	0.08 – 0.12	0.00	0.00 – 0.00	1.08	0.41 – 1.57
SERPINB2	0.00	0.00 – 0.02	0.11	0.08 – 0.15	0.00	0.00 – 1.00	2.14	1.61 – 2.71
SAMD10	0.06	0.04 – 0.09			0.82	0.29 – 1.08		
JAM3	0.00	0.00 – 0.02	0.10	0.08 – 0.13	0.00	0.00 – 0.00	2.11	1.31 – 2.51
ANKRD36C	0.06	0.04 – 0.07			0.80	0.02 – 1.35		
KIAA0319L	0.07	0.03 – 0.08			1.00	0.03 – 1.62		
ANP32B	0.00	0.00 – 0.01	0.09	0.06 – 0.11	0.00	0.00 – 0.00	1.13	0.33 – 1.93
HIST1H1A	0.05	0.03 – 0.09			1.13	0.41 – 1.70		
HLA-B	0.00	0.00 – 0.02	0.08	0.06 – 0.12	0.00	0.00 – 0.00	1.33	0.57 – 1.84
PEBP1	0.05	0.03 – 0.06			1.00	0.34 – 1.29		
SYT5	0.04	0.03 – 0.06			1.08	0.07 – 1.60		
EP300	0.00	0.00 – 0.02	0.07	0.05 – 0.10	0.00	0.00 – 0.00	0.73	-0.01 – 1.31
DMBT1	0.00	0.00 – 0.00	0.07	0.05 – 0.09	0.00	0.00 – 0.00	0.85	-0.14 – 1.46
LAMP1	0.00	0.00 – 0.01	0.06	0.04 – 0.08	0.00	0.00 – 0.00	1.43	0.91 – 2.00
MUC4	0.00	0.00 – 0.01	0.07	0.04 – 0.10	0.00	0.00 – 0.95	1.15	0.53 – 1.67
TLR5	0.00	0.00 – 0.00	0.06	0.04 – 0.08	0.00	0.00 – 0.00	0.99	0.10 – 1.67
TNC	0.00	0.00 – 0.01	0.06	0.04 – 0.09	0.00	0.00 – 0.00	0.50	-0.26 – 1.05
CD164	0.00	0.00 – 0.01	0.06	0.04 – 0.08	0.00	0.00 – 0.00	0.71	-0.63 – 1.24
CEACAM1	0.00	0.00 – 0.01	0.05	0.04 – 0.07	0.00	0.00 – 0.00	1.20	0.78 – 1.78
MUC1	0.00	0.00 – 0.01	0.06	0.04 – 0.09	0.00	0.00 – 0.00	0.87	0.30 – 1.68
B2M	0.00	0.00 – 0.00	0.05	0.04 – 0.06	0.00	0.00 – 0.00	0.93	0.18 – 1.69

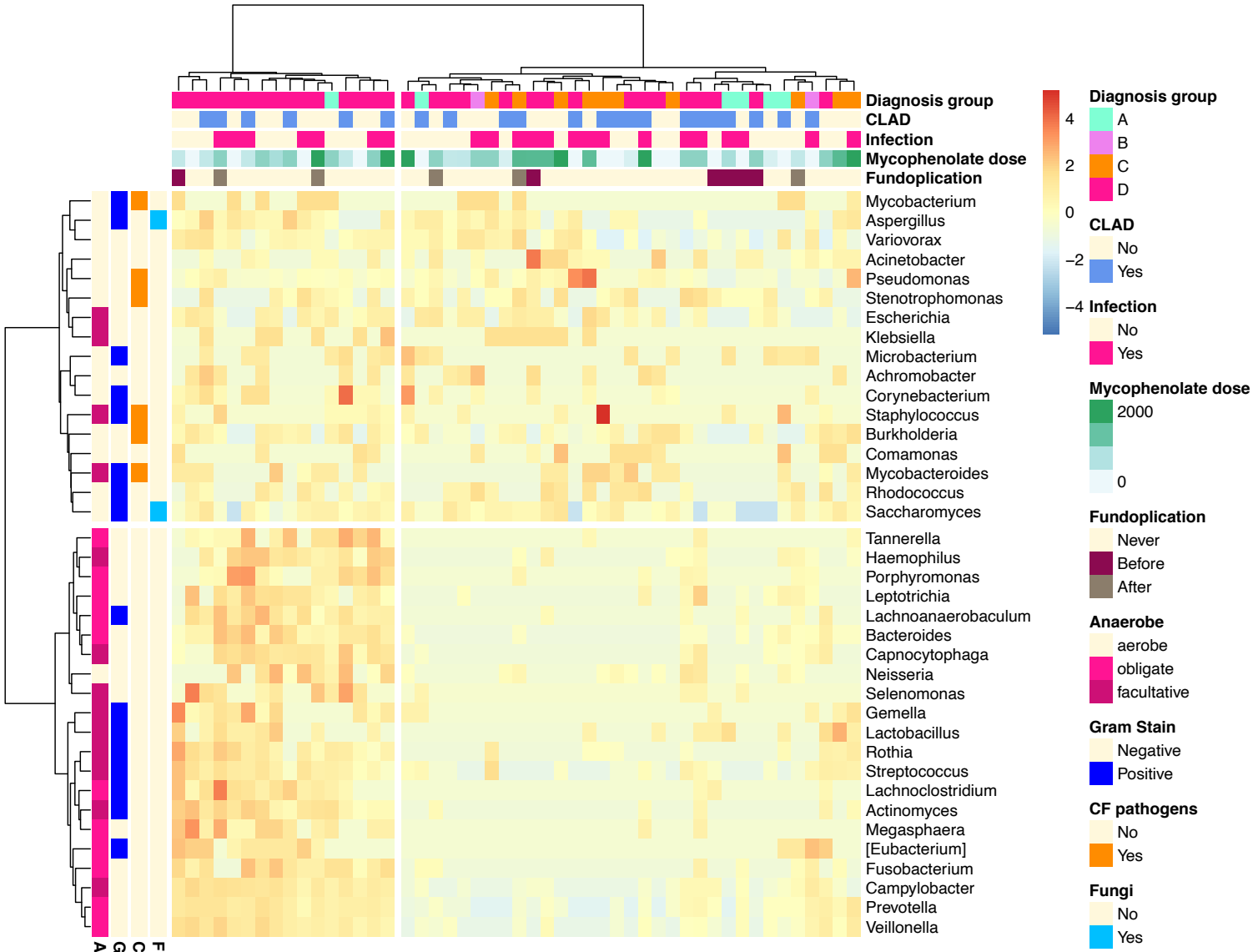
Leave-one-out cross validation was performed yielding 49 random forest models for classifying CLAD versus normal. Random forest models used either all available gene expression data (all genes) or the gene expression data limited to the set of genes also measured in tissue by nanoString digital RNA counting. Variable importance was extracted from each and shown here as median with interquartile range. Mean standard error is computed as the difference of prediction error for the “out-of-bag” data for each tree minus the error after permuting the predictor variables, normalized by standard deviation. Node purity increase is calculated from the residual sum of the squares. The top 50 genes are shown, while 93 genes had a median MSE greater than 0.

Supplemental Table 3: Bacterial, fungal, and viral pathogens identified in the no CLAD, infection and CLAD, infection groups

Species	Stable	CLAD
Aspergillus niger	5	3
Haemophilus parainfluenzae	3	3
Pseudomonas aeruginosa	3	2
Aspergillus fumigatus	3	1
Rhinovirus	1	1
Staphylococcus aureus	1	1
Aspergillus nidulans	0	1
Enterobacter	1	0
Mucorales	1	0
Serratia marcescens	0	1
Stenotrophomonas maltophilia	0	1

In the infection groups, there was a median of one pathogen per subject, with a range 1–4.

Supplemental Figure 2: Heatmap of commonly observed genera.



Bacteria and fungi identified by metagenomic sequencing in ≥25% of samples are shown on the y-axis, with airway brushes on the x-axis. No viruses were consistently identified. Subjects are annotated by fundoplication status, most recent daily mycophenolate dose, lung transplant indication diagnosis group, and status of infection and CLAD. Species are annotated as fungi, typical cystic fibrosis (CF) pathogens, by gram stain positivity, and oxidative metabolism type. One major cluster included predominantly group D (fibrosis) lung transplant recipients, for whom anaerobic bacteria were prevalent.

Supplemental Methods:

Cohort details: Immunosuppression included prednisone at 0.1 mg/kg per day, tacrolimus dosed to a trough concentration of 8–10 ng/ml, azithromycin 250 mg three times per week, and daily omeprazole. Mycophenolic acid was targeted 2 g/day, but was held or adjusted for leukopenia, skin cancer, or other serious side effects.

Matching: After two clinician review to confirm the diagnosis of CLAD, all available CLAD cases were included and matched to CLAD-free controls. The pool of potential CLAD-free controls was defined by the absence of CLAD diagnosis for at least 1 year after the airway brushing event. Cases were matched on age and culture status to the best available control by minimizing a mismatch penalty function of [0.25 * log (difference in post-transplant weeks) + number of pathogenic genera differing between BAL cultures]. For example, if a CLAD case was included at 24 months post-transplantation without infection, we targeted a subject with a non-CLAD subject at 24 months post-transplantation without BAL evidence of infection. For both cases and controls, airway brushing could have been for-cause (symptomatic) or for routine surveillance (asymptomatic) per institutional protocol (Supplemental Table 1). Of note, some controls subsequently went on to develop CLAD and one control subject was retransplanted at 152 days for acute respiratory distress syndrome with no evidence of CLAD. No subject was included more than once.

Subject characteristics and histopathology findings (ISHLT rejection grades (7), presence of macrophages and bronchial-associated lymphoid tissue, and large-airway inflammation or lymphocytic bronchitis (as assessed by or E-grade) were compared by Student's t-test or chi^2-test for continuous and categorical variables, respectively. Similar findings were obtained with a linear and logistic regression models with pathology scores as the dependent variable and CLAD and infection status as predictor variables.

Statistical power: Our primary hypothesis was that LB metagene scores would differ between CLAD cases and controls. With 27 controls and 22 cases, we had 80% power to detect an effect size (d) of 0.82.

Gene expression analysis: RNA sequencing gene counts were normalized and differential gene expression determined using DESeq2. Digital RNA counts were normalized using the "NanoStringNorm" package, as previously described (8). Global differences in gene expression were determined by PERMANOVA and visualized by multidimensional scaling plot. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses was performed on differentially expressed genes (alpha = 0.01) using "goseq". CLAD-associated gene expression differences in brushings and transbronchial biopsies were determined by unpaired t-tests on log-normalized count data and combined scores were determined using Stouffer's method. The Benjamini–Hochberg method was used to false discovery rate (FDR)-adjust P-values with a cutoff FDR of ≤ 0.1 considered acceptable.

Metagene scores were determined as the sum of normalized gene counts for a given gene set, normalized to a mean of 0 and a standard deviation of 1. Published gene sets included MSigDB Hallmark pathways (9) and genes upregulated in lymphocytic bronchitis (8). Differences between metagene scores between groups were calculated using Wilcoxon rank sum tests.

To illustrate rate of decline in CLAD and changes in MSigDB metagene scores over time, values versus time from brush were plotted using locally estimated scatterplot smoothing (LOESS) regression.

Graft survival: Time to graft failure was defined as date of retransplant or death minus date of airway brush, censored at most recent spirometry date. Kaplan–Meier models were plotted using "survminer," with log-rank p-value shown. Cox proportional hazards models were performed using "survival" and included LB metagene score as a continuous variable, CLAD status, single versus double lung transplant, UNOS diagnosis group (with group D as referent), recipient age and gender. Models were compared by chi-square test for the analysis of deviance table (anova).

Microbiome: Metagenomic identification of microbial pathogens at the genus level was performed using IDSeq (10). A negative control water sample was sequenced alongside clinical samples and control reads were subtracted from clinical samples. Non-fungal eukaryotes and microbes found as common contaminants of metagenomic sequence preps were excluded, as listed below. Alpha diversity was calculated with Simpson and Shannon metrics. Association between alpha diversity metrics and LB metagene scores were assessed by linear regression models. Beta-diversity was assessed by Bray–Curtis dissimilarity, although similar results were obtained when UniFraq phylogenetic distances were calculated by alignment to the Human Oral Microbiome (11). PERMANOVA was performed using "vegan" and diversity analyses used "phyloseq" and "DEseq2" R packages.

Machine learning models: To compare gene expression datasets for classification of CLAD status, we built lasso-penalized logistic regression and random forest models, using the "glmnet" and "randomForest" packages, respectively. These models exclusively included gene expression data and were limited to the set of genes present in both nanoString and RNAseq datasets. To prevent overfitting, we used leave-one-out cross validation with random selection of controls to match the number of cases. Thus, there was one machine learning model of each type generated for each data point. Feature importance was extracted from each and median values are shown in Supplemental Table 2. Classifier accuracy was assessed by area under the receiver operating curve (ROC), with 95% confidence intervals (CI) and between curve comparisons done using DeLong's methods.

Excluded genera: The following genera were excluded based on high likelihood of being metagenomic sequencing or sample collection reagent contaminants: *Acidovorax, Agromyces, Aquabacterium, Azospirillum, Bodo, Bosea, Bradyrhizobium, Brassica, Candidatus, Caulobacter, Cedrus, Chlorella, Chroococcidiopsis, Chryseobacterium, Chrysolepis, Clostridium, Cocconeis, Coelastrum, Cryptomonas, Cyclidium, Cyclotella, Deinococcus, Delftia, Dysteria, embryophyte, Entosiphon, Flavobacterium, Gemmata, Glaesserella, Glutamicibacter, Hariotina, Herbinix, Hydrodictyon, Hydrogenophaga, Hymenobacter, Kadipiro, Leifsonia, Leishmania, Limnohabitans, Mesorhizobium, Mesorhizobium, Methylorubrum, Mitella, Naegleria, Neobodo, Nitrobacter, Nitzschia, Oligotropha, Paenibacillus, Paraburkholderia, Paracoccus, Paracoccus, Parastrongyloides, Paucibacter, Pectinodesmus, Pediastrum, Picea, Pinus, Pirellula, planctomycete, Polynucleobacter, Prosthecobacter, Quercus, Ramlibacter, Rhizobacter, Rhizobium, Rhodobacter,*

Rhodoferax, Rhodopseudomonas, Saccharopolyspora, Saccharum, Salpingoeca, Sinorhizobium, Solanum, Sphaeroeca, Sphingobium, Sphingopyxis, Stentor, Taxus, Teleaulax, Tetradesmus, Tetrahymena, Trypanosoma, Vigna, Xanthomonas, and *Ziziphus*.

Supplemental References

1. Peng T, Frank DB, Kadzik RS, Morley MP, Rathi KS, Wang T et al. Hedgehog actively maintains adult lung quiescence and regulates repair and regeneration. Nature 2015;526(7574):578-582.
2. Tsao PN, Matsuoka C, Wei SC, Sato A, Sato S, Hasegawa K et al. Epithelial Notch signaling regulates lung alveolar morphogenesis and airway epithelial integrity. Proc Natl Acad Sci U S A 2016;113(29):8242-8247.
3. Schmid A, Sailland J, Novak L, Baumlin N, Fregien N, Salathe M. Modulation of Wnt signaling is essential for the differentiation of ciliated epithelial cells in human airways. FEBS Lett 2017;591(21):3493-3506.
4. Pasnupneti S, Nicolls MR. Airway hypoxia in lung transplantation. Curr Opin Physiol 2019;7:21-26.
5. Faust HE, Golden JA, Rajalingam R, Wang AS, Green G, Hays SR et al. Short lung transplant donor telomere length is associated with decreased CLAD-free survival. Thorax 2017;72(11):1052-1054.
6. Gottlieb J, Neurohr C, Muller-Quernheim J, Wirtz H, Sill B, Wilkens H et al. A randomized trial of everolimus-based quadruple therapy vs standard triple therapy early after lung transplantation. Am J Transplant 2019;19(6):1759-1769.
7. Stewart S, Fishbein MC, Snell GI, Berry GJ, Boehler A, Burke MM et al. Revision of the 1996 working formulation for the standardization of nomenclature in the diagnosis of lung rejection. J Heart Lung Transplant 2007;26(12):1229-1242.
8. Greenland JR, Wang P, Brotman JJ, Ahuja R, Chong TA, Kleinhenz ME et al. Gene signatures common to allograft rejection are associated with lymphocytic bronchitis. Clin Transplant 2019;33(5):e13515.
9. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1(6):417-425.
10. Ramesh A, Nakielny S, Hsu J, Kyohere M, Byaruhanga O, de Bourcy C et al. Metagenomic next-generation sequencing of samples from pediatric febrile illness in Tororo, Uganda. PLoS One 2019;14(6):e0218318.
11. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH et al. The human oral microbiome. J Bacteriol 2010;192(19):5002-5017.