It is made available under a CC-BY-ND 4.0 International license .

1 Bronchoalveolar lavage metabolome dynamics reflect underlying disease and chronic

2 lung allograft dysfunction

3

4 Christian Martin ^a, Kathleen S. Mahan ^b, Talia D. Wiggen ^c, Adam J. Gilbertsen ^c, Marshall I.

5 Hertz^b, Ryan C. Hunter^{c*} and Robert A. Quinn^{a*}

^a Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing,
 Michigan USA

7 Michigan, USA.

8 ^b Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Department of Medicine,

9 University of Minnesota, Minneapolis, Minnesota, USA.

^c Department of Microbiology and Immunology, University of Minnesota Medical School,
 Minneapolis, Minnesota, USA.

¹² Denotes co-corresponding authors: Ryan C. Hunter, Department of Microbiology and ¹³ Immunology, University of Minnesota Medical School, 689 23rd Ave SE, Minneapolis, ¹⁴ Minnesota, 55455, <u>rchunter@umn.edu</u>; Robert A. Quinn, 603 Wilson Rd. Rm 120, Department ¹⁵ of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, ¹⁶ 48824, <u>quinnrob@msu.edu</u>

17

18 Abstract

Background Progression of chronic lung disease often leads to the requirement for a lung transplant (LTX). Despite improvements in short-term survival after LTX, chronic lung allograft dysfunction (CLAD) remains a critical challenge for long-term survival. This study investigates the relationship between the metabolome of bronchoalveolar lavage fluid (BALF) from subjects post-LTX with underlying lung disease and CLAD severity.

Methods Untargeted LC-MS/MS metabolomics was performed on 960 BALF samples collected over 10 years from LTX recipients with alpha-1-antitrypsin disease (AATD, n=22), cystic fibrosis (CF, n=46), chronic obstructive pulmonary disease (COPD, n = 79) or pulmonary fibrosis (PF, n=47). Datasets were analyzed using machine learning and multivariate statistics for associations with underlying disease and final CLAD severity.

Results. BALF metabolomes varied by underlying disease state, with AATD LT recipients being particularly distinctive (PERMANOVA, p=0.001). We also found a significant association with the final CLAD severity score (PERMANOVA, p=0.001), especially those with underlying CF. Association with CLAD severity was driven by changes in phosphoethanolamine (PE) and phosphocholine lipids that increased and decreased, respectively, and metabolites from the

34 bacterial pathogen Pseudomonas aeruginosa. P. aeruginosa siderophores, quorum-sensing

It is made available under a CC-BY-ND 4.0 International license .

35 quinolones, and phenazines were detected in BALF, and 4-hydroxy-2-heptylquinoline (HHQ) 36 was predictive of the final CLAD stage in samples from CF patients (R=0.34; $p\leq0.01$). 37 Relationships between CLAD stage and P. aeruginosa metabolites were especially strong in 38 those with CF, where 61% of subjects had at least one of these metabolites in their first BALF 39 sample after transplant. 40 **Conclusions:** BALF metabolomes after LTX are distinctive based on the underlying disease 41 and reflect final CLAD stage. In those with more severe outcomes, there is a lipid transition from 42 PC to predominantly PE phospholipids. The association of P. aeruginosa metabolites with 43 CLAD stages in LTX recipients with CF indicates this bacterium and its metabolites may be 44 drivers of allograft dysfunction.

45

Key messages: Despite the high prevalence of CLAD among LTX recipients, its pathology is not well understood, and no single molecular indicator is known to predict disease onset. Our machine learning metabolomic-based approach allowed us to identify patterns associated with a shift in the lipid metabolism and bacterial metabolites predicting CLAD onset in CF. This study provides a better understanding about the progression of allograft dysfunction through the molecular transitions within the transplanted lung from the host and bacterial pathogens.

52

Key words Metabolomics; cystic fibrosis; bronchioalveolar lavage fluids; chronic lung allograft
dysfunction, chronic lung diseases; phosphoethanolamine; phosphocholine; quinolones, *Pseudomonas aeruginosa.*

56

57 Word count: 3,452

58

59

It is made available under a CC-BY-ND 4.0 International license .

61 Introduction

62 Lung transplantation (LTX) is a therapeutic option for patients who develop progressive 63 and severe chronic lung disease [1]. However, chronic lung allograft dysfunction (CLAD) 64 remains a major barrier to long-term LTX patient survival, as it affects 50% of patients at five years post-transplantation [2,3]. Bronchiolitis obliterans syndrome (BOS), a crucial phenotypic 65 66 manifestation of allograft dysfunction [4,5], is used as a scaled measure of disease progression 67 (i.e., CLAD stage) [6]. Notoriously a heterogenous disease, BOS severity scores were recently 68 updated in 2019 [6] to better characterize the stages of severity experienced by LTX recipients. 69 These CLAD stages can occur post-LTX in patients with different pre-LTX lung diseases such 70 as alpha-1-antitrypsin deficiency disease (AATD), cystic fibrosis (CF), chronic obstructive 71 pulmonary disease (COPD), and pulmonary fibrosis (PF) [7-10].

Bronchoalveolar lavage fluid (BALF) is a valuable sample type for studying pathological characteristics that develop after LTX due to the direct proximity of BALF sample origin to the site of stress and/or injury in the lung [11]. Metabolites detected within BALF can be indicative of physiological processes of disease progression, microbial lung burden, and possibly, allograft rejection [12,13]. However, there are few studies of BALF metabolomes describing the composition, microbial virulence factors, amino acids, lipids, and a myriad of pharmaceuticals [14–17].

79 Untargeted mass spectrometry (MS)-based metabolomics allows exploration of the 80 chemical diversity associated with a wide range of biological samples [18,19]. However, 81 navigating the diverse chemical data generated in untargeted metabolomics studies, in which a 82 large proportion of detectable metabolites are unknown, remains challenging [18,19]. Advances 83 in bioinformatic analyses of MS data have enabled a more comprehensive interpretation of 84 biological information contained within these highly technical and large datasets. For example, 85 the Global Natural Products Social molecular networking web platform (GNPS) has simplified 86 the exploration of the chemical space of metabolomes. GNPS is a tandem mass spectrometry 87 (MS/MS) automated data organizational tool that enables comparisons of MS/MS fragmentation patterns among samples to cluster and visualize related molecules in a spectral network [20.21]. 88 GNPS can be paired with metabolite feature quantification algorithms to create a 89 90 comprehensive workflow for untargeted metabolomics of clinical samples [20,22]. In this study, we applied this approach to 960 BALF samples collected longitudinally from 194 LTX recipients 91 92 to determine whether post-transplant BALF metabolomes reflect the underlying disease 93 diagnosis; and whether specific molecular signatures correlate with allograft dysfunction.

It is made available under a CC-BY-ND 4.0 International license .

95 Materials and Methods

96 Study Design, Subjects, and BOS-Grading.

Subjects with AATD, CF, COPD, and IPF who underwent lung transplantation at the 97 98 University of Minnesota consented to have a portion of their post-LTX BALF used for research. 99 BALF was collected over a 10-year period from 2002 to 2012 as a routine procedure per 100 transplant protocol or if undergoing diagnostic BALF collection due to clinical indications, such 101 as new radiographical changes, new respiratory symptoms, or a decrease in FEV₁. Inclusion 102 criteria were \geq 18 years of age: diagnosis of AATD, CF, COPD, or PF; and receipt of a bilateral 103 or single lung transplant. Exclusion criteria were < 18 years of age and the inability to provide 104 consent and/or tolerate the BALF procedure. Clinical characteristics of subjects in this study, 105 including disease progression scores and samples collected for each, are displayed in Table 1 106 and Table S1. BOS-grade was recorded for patients based on the changes in their forced 107 expiratory volume in one second (FEV₁), initially determined using definitions of the International 108 Society for Heart and Lung Transplantation (ISHLT) [23], and updated here to reflect current 109 CLAD definitions [6]. The final CLAD stage was known for each patient, while some measures 110 were determined after sample collection ceased. This protocol was approved by the University 111 of Minnesota IRB (STUDY00004547).

112

113 Bronchoalveolar lavage sampling material.

Bronchoalveolar lavage fluid (BALF) was collected via orotracheal or nasotracheal bronchoscopy. Sterile saline was instilled into a subsegmental bronchus after advancement and occlusion of the airway lumen. BALF was separated into ~1.5 mL aliquots and frozen at -80 °C for further untargeted metabolomic analysis (see supplementary methods). All 960 samples were included for analysis of variation (ANOVA) among underlying disease types and associations with final CLAD stages.

120

121 Organic extraction and LC-MS/MS analysis

122 Methanolic extracts of 960 BALF samples (50:50 v/v) were analyzed on a Thermo 123 QExactive^{IM} mass spectrometer coupled to a Vanguish Ultra-High-Performance Liguid 124 Chromatography (UHPLC) system (ThermoFisher) (see supplementary methods) [21,24]. 125 Subsequently, feature-based molecular networking (FBMN) was performed with a parent and 126 fragment mass ion tolerance of 0.02 Da, a cosine score of 0.65, and a minimum matched peaks 127 of minimum 4 [22]. FBMN is publicly available at 128 https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=756456ac794d48b0bba80dbe28e0de66,

It is made available under a CC-BY-ND 4.0 International license .

and raw data files are available in the MASSIVE data repository as <u>MSV000085760</u>. Data curation consisted of removing metabolites detected in blanks and removal of known pharmaceuticals and their related nodes in the molecular network (see supplementary methods, Figure S1, and Table S2). Presence/absence frequency and rank abundance of all molecular features were calculated after creating a cutoff by summing the relative abundances of features above 10 x e⁵ abundance in samples obtained from subjects with AATD, CF, COPD, or IPF.

135

136 Statistical

analysis

137 Variations in metabolomic data associated with underlying disease type and the final 138 CLAD stage for each subject were assessed using beta-diversity metrics and tested for 139 significance with permutational multivariate analysis of variance (PERMANOVA). Metabolome 140 variation was further assessed using random forest (RF) classification and regression analysis 141 (numerical CLAD stages, 0 to 4). A Bray-Curtis dissimilarity matrix was calculated on the entire 142 metabolome and used to generate principal coordinate analysis (PCoA) plots through the in-143 house tool ClusterApp and visualized in EMPeror [25]. PERMANOVA tests were performed for 144 diseases and CLAD stages (as a categorical variable) with subject-source as an interacting 145 factor to account for variation in the number of samples per subject. Post-hoc tests among 146 disease types were performed with R packages Devtools and Vegan (pairwise-adonis) [26,27]. 147 Variable importance plots from both RF approaches were used to identify metabolites driving 148 the variation observed. Pearson correlations were used to determine relationships between 149 metabolite abundance and final CLAD stages, as well as the correlation of P. aeruginosa 150 metabolites with its relative abundance determined using 16S rRNA sequence data from the 151 same samples. 16S rRNA sequence data were generated using bacterial genomic DNA 152 extracted from each sample, sequencing the V4 region using Illumina MiSeq TruSeq 2x300 153 paired-end technology, and sequence analysis in R as previously described [28]. The relative 154 abundance of *P. aeruginosa* was determined using DADA2 [29]. R packages random forest, 155 vegan and ggplot2 were also used for these analyses [30–32] (see supplementary methods).

156

157 Results

158 Sample collection and clinical design.

The two primary objectives for this study were: i) to determine if the metabolome of BALF collected after LTX was associated with the underlying pre-LTX lung diagnosis, and ii) if the data were associated with measures of CLAD severity across all subjects and within each underlying disease type. The dataset comprised longitudinal BALF samples (n = 960) collected

It is made available under a CC-BY-ND 4.0 International license .

163 over a 10-year period from subjects (n = 194) with one of four underlying conditions prior to 164 transplant; AATD (n = 22), CF (n = 46), COPD (n = 79), and PF (n = 47). Clinical parameters 165 and patient demographic information are presented in Table 1. Subjects developed CLAD at 166 different times and to varying degrees during the collection period. The final BOS-grade of all 167 subjects was known, even when this measure was recorded after BALF collection had ceased. 168 Thus, metabolome data variation was tested against this final BOS-grade measure.

169

170 BALF metabolome variation based on the underlying lung disease type pre-transplant

171 Since BALF samples were obtained via bronchoscopy, numerous drugs and xenobiotics 172 were identified in the dataset. We reasoned that these molecules could confound underlying 173 biologically significant trends, so they were removed from the metabolome data using Global Natural Product Social (GNPS) library searching and molecular networking to identify known 174 175 and chemically related pharmaceuticals in the GNPS libraries. After filtering pharmaceuticals 176 and sample contaminants (detected in controls), the entire BALF data set included 4755 177 molecular features, of which 361 had a spectral match to known compounds in GNPS libraries 178 [21].

179 PCoA ordination of the entire dataset, colored by pre-transplant disease diagnosis, 180 revealed significant differences between AATD, CF, COPD, and PF cohorts (PERMANOVA F =181 3.91, p = 0.001, Figure 1a). Additionally, post-hoc testing showed significant pairwise 182 differences between diseases (Table 2), suggesting that allograft metabolomes are dependent, 183 in part, on the underlying disease of the transplant recipient. This unique disease signature was 184 especially marked in subjects with AATD, as evidenced by its low random forest (RF) 185 classification error (13.3%, Table S3) and separation by PCoA clustering (Figure 1a). We then 186 used supervised RF classification analysis to identify metabolites that most strongly 187 distinguished BALF samples by the underlying disease. The top 10 metabolites driving 188 differences between groups included phenylalanine, phosphocholines, and other lipids. Notably, 189 phenylalanine was particularly abundant in subjects with underlying AATD (pairwise p-value CF $p = 2.9 e^{-11}$, COPD $p = 4.7 e^{-14}$, and IPF $p = 3.1 e^{-13}$) (Figure 1b, Figure S2). 190

To further understand the nature of the unique BALF metabolome in LTX recipients with AATD, we analyzed metabolite presence/absence to determine the degree of chemical sharing among underlying diseases. Although most metabolites were shared across all four diseases, AATD unexpectedly had fewer molecules shared with the other diseases (either one other, or two others, Figure 1c). Rank abundance curves of the metabolome revealed that AATD had

It is made available under a CC-BY-ND 4.0 International license .

196 fewer overall metabolites, and these were less abundant than the other three disease types197 (Figure 1a,d).

198

199 Longitudinal allograft metabolomes are associated with CLAD severity outcomes

200 Each subject's final CLAD stage was then used to assess the relationship between the 201 entire longitudinal metabolome dataset and disease severity (n = 960). PERMANOVA 202 (categorical CLAD stage, F = 4.023, p = 0.001) and RF regression (linearized CLAD stage, 203 variance explained = 11.08%) revealed an overall association of the final CLAD stage with 204 collective longitudinal metabolome composition (Figure 2a). When tested separately on each 205 pre-LTX disease diagnosis, CLAD stages maintained their categorical significance for all but IPF 206 (Figure 2b). RF regression showed that LTX recipients with underlying CF had the strongest 207 association of their BALF metabolome variation with numerical CLAD stages (% variance 208 explained = 34.04), followed by COPD (14.36) and markedly less variation in IPF (4.65) and 209 AATD (3.16) (Figure S3). These results demonstrate a significant relationship between the 210 BALF metabolome after LTX and the final CLAD stage, which was particularly strong among 211 individuals with CF.

212

Altered phospholipids and *P. aeruginosa*-derived molecules drive association with the final CLAD stage

215 The association between CLAD stage and longitudinal metabolome composition, 216 particularly among CF subjects, motivated further analyses of specific metabolites that most 217 strongly influenced trends in the dataset. Variable importance plots from the RF analysis of the 218 entire dataset and each disease analyzed separately revealed several metabolites of interest 219 (Figure S3, S4). Molecular networking was then used to help annotate and identify these and 220 related molecules. With this approach, we identified four clusters ('molecular families') of 221 interest, including phosphoethanolamine (molecular family I), phosphocholine (molecular 222 families II and III), and quinolone-like-molecules (molecular family IV, Figure S5). The molecular 223 family I included the phospholipids lysophosphoethanolamine (lysoPE) 18:0/0:0 (m/z 482.3225 224 $[M+H]^+$, $C_{23}H_{49}NO_7P$, 3.4 ppm error) and lysoPE 18:1/0:0 (m/z 480.3073 $[M+H]^+$, $C_{23}H_{47}NO_7P$, 225 2.40 ppm error). Molecular family II included phosphatidylcholine 16:0/14:0 (m/z 728.5183 226 $[M+Na]^+$, $C_{38}H_{76}NO_8P$, 2.42 ppm error). Molecular family III corresponds to PC 16:0/18:0 (m/z227 780.5529 [M+Na]⁺, C₄₂H₈₀NO₈P -2.20 ppm error). Finally, in molecular family IV, we identified 4-228 hydroxy-2-heptylquinolone (HHQ, m/z 244.1689 [M+H]⁺, C₁₆H₂₂NO 2.8 ppm error), a 229 Pseudomonas aeruginosa-derived quinolone and known quorum sensing metabolite that plays

an integral role in the gene expression and physiology of this opportunistic pathogen (FiguresS5 and S6) [33].

232 We then plotted the relationship between the feature abundances of each metabolite of 233 interest with the final CLAD stages across the entire dataset. Both lysoPE(18:0/0:0) and 234 lysoPE(18:1/0:0) significantly increased with disease progression (Pearson r = 0.23, p = 6.9e-13235 , and r = 0.16, p = 0.017, respectively). In contrast, phosphocholine molecules PC (16:0/14:0) 236 and PC(16:0/18:0) decreased as final disease severity increased (Pearson r = -0.22, p = 4.4e-237 10 and r = -0.19, p = 1.9e-08, respectively). Linear regression analysis of HHQ abundance and 238 individual subjects' final CLAD stage also showed a significant increase as allograft dysfunction 239 worsened (Pearson r = 0.14, p = 0.0035) (Figure 3a).

240 Finally, we parsed the molecular dynamics of these molecules based on the underlying 241 disease and identified unique trends within each condition. LvsoPE (18:0/0:0) significantly 242 increased with disease progression in CF (Pearson r = 0.33, $p \le 0.001$), followed by COPD (r =243 0.28, $p \le 0.001$) and IPF ($r = 0.14 \ p \le 0.001$), but not in AATD. LysoPE (18:1/0:0) abundance 244 was positively correlated with IPF only (Pearson r = 0.42, $p \le 0.001$). The feature abundance of 245 PC (16:0/14:0) significantly decreased as disease severity worsened only in CF and COPD (Pearson r = -0.41, $p \le 0.001$ and r = -0.28 $p \le 0.001$, respectively), whereas PC (16:0/18:0) 246 significantly decreased in CF, COPD and IPF (Pearson r = -0.27, $p \le 0.001$; r = -0.24, $p \le 0.001$; 247 248 and $r = -0.19 \ p \le 0.001$, respectively). Neither of these molecules varied with CLAD stages in 249 AATD. In the case of HHQ, its molecular abundance significantly increased with disease 250 progression in CF (Pearson r = 0.34, $p \le 0.001$) (Figure 3b) but not the others, indicating that 251 the trend seen in the complete dataset was driven by the CF samples.

252

253 Pseudomonas aeruginosa molecular signatures in subjects with CF after LTX

254 The association of HHQ with CLAD severity led to further analysis of the diversity of P. 255 aeruginosa metabolites and their relationship with clinical outcomes, particularly in those 256 subjects with CF. Molecular networking allowed us to identify additional molecular signatures 257 from the pathogen, including 2-heptyl-4-guinolone-N-oxide (HQNO; m/z 260.1645 [M+H]⁺, 258 C₁₆H₂₂NO₂; 10.67 ppm), 2-nonyl-4-quinolone (HNQ; *m/z* 270.1852 [M+H]⁺, C₁₇H₂₄NO; 7.45 259 ppm), pyocyanin (m/z 211.0866 [M+H]⁺, C₁₃H₁₁N₂O; 5.64 ppm) and pyochelin (m/z 325.0675 260 [M+H]⁺, C₁₄H₁₆N₂O₃S₂; 10.46 ppm). Molecular network node mapping based on disease showed 261 that the entire molecular family of quinolones and pyochelin were enriched in subjects with 262 underlying CF (pie charts within nodes) (Figure 4a). We were therefore interested in determining 263 when these molecules were detected in the longitudinal BALF samples of each subject. We

264 found that ~60% of samples (n = 46) from subjects with CF were positive for at least one P. 265 aeruginosa molecule in the first BALF sample collected (Figure 4b), and the abundances were 266 particularly high at this first-time point (Figure 4c). This proportion of positive samples stayed 267 relatively stable, with a slight decrease in successive BALFs collected (Figure 4b). Some 268 subjects showed acquisition of *P. aeruginosa* molecules as time since transplant progressed, 269 whereas most initially had high abundances and a subsequent decrease (Figure 4c). However, 270 a Pearson correlation between the abundance of these molecules and the time since LTX was 271 not significant (R = 0.074; p = 0.18). The abundance of these molecules did increase (R = 0.34; 272 p = 0.0046) with the relative abundance of *Pseudomonas* sp. (Figure 4d,e).

273

274 Discussion

275 In this study, we applied untargeted metabolomics to 960 BALF samples from patients 276 who had undergone LTX for chronic lung disease. We had a particular interest in metabolite 277 variation based on the underlying disease and the "gold standard" measure of disease severity 278 (i.e., CLAD stage). The post-LTX metabolome data differed based on the underlying disease, 279 with the unique profile found in AATD. Other chronic lung diseases (CF, COPD, PF) were 280 significantly different from one another but more difficult to distinguish overall. The uniqueness 281 of AATD was driven by differences in aromatic amino acids (phenylalanine), an overall lack of 282 shared molecules with the other diseases, and a lower abundance of metabolite features 283 overall. We note that all four underlying conditions have unique etiologies, and the findings here 284 indicate that, somewhat unexpectedly, the chemical environment of the lung allograft reflects 285 the initial disease of the recipient, particularly in the case of AATD. The high abundance of 286 phenylalanine in allografts from subjects with AATD may reflect increased proteolysis, which is 287 a hallmark of this disease [23], though further research is needed to confirm this hypothesis.

288 The overall annotation rate of MS/MS spectra in our dataset was 7.6%, which is not 289 uncommon in untargeted metabolomics studies [34]. Low levels of metabolite identification are a 290 known challenge in metabolomics and can limit the ability to infer mechanistic associations with 291 disease severity, although annotation rates in untargeted metabolomics experiments are 292 increasing as new search algorithms and databases become available [22,35-37]. 293 Nevertheless, there is considerable power in using comprehensive molecular data from 294 untargeted metabolomics experiments to identify biological trends, even when the molecular 295 structures are not known. An important step for this approach, which we performed here, is the 296 removal of pharmaceuticals and xenobiotics, which can be highly abundant and overwhelm 297 underlying biological signals. The molecular networking algorithm applied here greatly increases

the ability to 'clean up' the metabolome by enabling the identification and visualization of pharmaceuticals and their chemical relatives.

300 Bronchoalveolar lavage fluids have been used recently for describing molecular changes 301 in lipids and metabolites after LTX [16]; however, the association of these compounds with 302 allograft dysfunction has not been addressed. One of the strongest signatures in our 303 metabolomic data was the association of the metabolome with the final CLAD stage of each 304 subject, and this was particularly strong for those with CF. In this case, many of the molecular 305 features driving the association had annotations in the GNPS libraries, and they were primarily 306 phospholipids and bacterial metabolites. A closer analysis of these molecules identified a 307 transition in lipid species associated with an increased final CLAD stage. Subjects whose final 308 CLAD severity was lower had more abundant phosphocholine lipids in their longitudinal BALF 309 metabolome, while those with more severe CLAD outcomes had a higher abundance of phosphoethanolamine lipids. This phospholipid lipid transition may indicate a disruption in 310 311 airway surface liquid (ASL) in the transplanted lung as the disease progresses or alterations in 312 lipid metabolism. ASL is mainly composed of phospholipids (90%) and proteins (10%), which 313 are products of surface and submucosal gland epithelia and resident phagocytic cells [38]. 314 Approximately 70–80% of ASL lipids are dipalmitoyl phosphatidylcholine (DPPC) [39,40] and PE 315 is a major phospholipid in lung surfactant. The increase in PE abundance associated with CLAD 316 severity found here may be due to the influx of neutrophils or other inflammatory cells containing 317 this lipid. Another potential source is allograft-colonizing microbiota, as PE lipids are known to 318 be a component of bacterial membranes [41,42].

319 Detection of quorum-sensing metabolites from P. aeruginosa in BALF and the 320 association of HHQ with final CLAD severity implicates the bacterium and its metabolism in 321 disease progression. These molecules were particularly prevalent in subjects with underlying 322 CF, where this bacterium is a known opportunistic pathogen responsible for chronic airway 323 infection [43–45] [43]. P. aeruginosa regulates the production of its virulence factors through its 324 quorum sensing system, which plays an important role in CF pathogenesis. Multiple quinolones 325 from P. aeruginosa such as HQNO, HNQ, and PQS (Pseudomonas quinolone signal), were 326 detected in BALF, as were other two other small molecule virulence factors pyocyanin and 327 pyochelin. Quorum sensing is mediated through N-acyl-homoserine lactones (AHLs) and alkyl 328 quinolones (AQ) [46], the latter of which were detected in this study. The correlation between 329 the abundance of these molecules and the final CLAD stage indicates they could potentially be 330 explored as biomarkers of bacterial infection in CF patients post-LTX and as indicators of CLAD 331 progression. Furthermore, they are relatively easy to detect with LC-MS/MS rapidly after sample

It is made available under a CC-BY-ND 4.0 International license .

extraction. Finally, we note that many of these *P. aeruginosa* small molecules were detected in the first BALF sample after LTX. This observation supports the hypothesis that the bacterium readily re-colonizes the respiratory tract post-LTX from its reservoir in the upper airways (i.e., paranasal sinuses) [44]. Though data from this study cannot directly assess this infection reservoir, the detection of quinolones in BALF so early after LTX is an important finding for understanding the pathogenesis of CLAD.

338

339 Conclusions

340 The BALF metabolome after lung transplant revealed differences based on underlying lung 341 disease type and association with final CLAD severity. An important metabolic trend was a shift 342 in the relative abundance of phosphocholine and phosphoethanolamine lipids that were 343 predictive of a subject's final CLAD stages with CF. These findings indicate potential predictive 344 value for the lipid profile as an indicator of disrupted airway surface liquid and impending CLAD 345 severity. Importantly, our LC-MS/MS approach readily detected virulence metabolites from the 346 bacterial pathogen P. aeruginosa, especially in CF samples, which were associated with poor 347 CLAD outcomes. This study provides a picture of the molecular transitions within the 348 transplanted lung from the host and bacterial pathogens that may help understand the 349 progression of allograft dysfunction and merits further study with more targeted approaches for 350 the molecules identified.

It is made available under a CC-BY-ND 4.0 International license .

351 Acknowledgment

- 352 We thank Anthony Schilmiller for his constant support during sample processing in the mass
- 353 spectrometry and metabolomics core laboratory at Michigan State University. The CF
- 354 Foundation grant HUNTER18ABO and the National Institute of Allergy and Infectious Diseases
- 355 (R01AI145925) provided funding.

356 Author contributions:

- R. Hunter and R. Quinn conceptualized the study. K. Mahan performed BALF/CLAD stages tabulation. C. Martin, T. Wiggen, and R. Quinn performed data acquisition, analysis, and interpretation. C. Martin, R. Quinn, and R. Hunter wrote the original and final versions of the manuscript. K. Mahan and M. Hertz edited the manuscript. All authors read and approved the
- 361 final manuscript.

362 **Disclosure statement:**

- 363 All authors in the presented manuscript have no conflicts of interest to disclose.
- 364 Sources of funding:
- 365 Cystic Fibrosis Foundation (HUNTER18AB0) and National Institutes of Health (R01AI145925).
- 366

It is made available under a CC-BY-ND 4.0 International license .

Clinical parameters	AATD	CF	COPF	IPF
Number of subjects	22	46	79	47
Sex ratio (F/M)	0.69 (9/13)	1.20 (25/21)	1.20 (43/36)	0.42 (14/33)
Average of age at LTX (range)	55 (42 – 55)	36 (20 – 54)	59 (44 – 74)	58 (39 – 68)
Average of last recorded FEV1 (range)	64 (18 - 96)	69 (17 – 100)	65 (23 – 100)	68 (26 – 99)
Average of last recorded CLAD stage (0 to 4)	1.57	1.35	1.62	1.36
Number of subjects with P. aeruginosa (%)	11 (50%)	32 (70%)	7 (5%)	4 (9%)
BALF samples	130	185	362	283
BALF average/subject (range)	5.01 (2-12)	4.02 (1-9)	4.58 (2-9)	6.02 (3-10)
Average of sample collection (years)	2.93	3.33	3.64	4.05

367

Table 1. Clinical characteristics of subjects (total n=194) and samples (total n=960) collected post-LTX grouped by prior underlying disease – alpha-1 anti-trypsin deficiency (AATD), cystic fibrosis (CF), chronic obstructive pulmonary disease (COPD), and idiopathic pulmonary fibrosis (IPF). BALF sampling information, microbiology associated with *Pseudomonas aeruginosa*, age at transplant (LTX), sex ratio (female/male). The averages of the last recorded forced expiratory volume per 1 second expressed as a percent of predicted (%FEV₁) and the CLAD stage are displayed based on pre-transplant diagnoses.

375

376

It is made available under a CC-BY-ND 4.0 International license .

Pairwise Comparisons	F-value	p-value
COPD and IPF	26.5689	0.001
COPD and CF	2.1869	0.002
COPD and IPF	1.7698	0.017
AATD and CF	21.2079	0.001
AATD and IPF	28.6926	0.001
CF and IPF	3.1142	0.001

- 378 **Table 2** Post-hoc pairwise comparisons of BALF metabolome among different diseases (AATD,
- 379 CF, COPD, and IPF).

It is made available under a CC-BY-ND 4.0 International license .





Figure 1. Lung allograft metabolomic profiles vary by pre-LTX condition. a) PCoA plot of a 383 384 Bray-Curtis distance matrix of lung allograft BALF metabolomic profiles. Samples colored by 385 pre-LTX disease and demonstrate significant differences among disease types (PERMANOVA, 386 F=3.91, p = 0.001, random forest out-of-bag (OOB) error = 41.03%). b) Feature abundance of 387 DL-Phenylalanine by disease state. Kruskal-Wallis test and post-hoc Dunn test p-values are 388 shown. c) Uniqueness and sharing of BALF metabolites across diseases. Molecule presence or 389 absence was determined and plotted by the number of molecules that are unique, shared with 390 one other, two others, or all four diseases. d) Ranking molecular abundance curves of BALF 391 metabolome colored by disease.

392

It is made available under a CC-BY-ND 4.0 International license .



Figure 2. Principal coordinate analysis plots of Bray-Curtis distance matrices calculated for BALF metabolome based on CLAD stage measurements: Statistics from the categorical PERMANOVA testing (F and p-value) and linear variation based on the RF analysis (%Var) are shown for each plot. a) PCoA plot of the entire BALF metabolome colored by the final CLAD stages. b) PCoA plots of the entire BALF metabolomic data separated by disease type colored by the final CLAD stages.

It is made available under a CC-BY-ND 4.0 International license .



401 **Figure 3** a) Scatter plots and linear regression analysis of feature abundances of target 402 molecules lysoPE (18:0/0:0), lysoPE (18:1/0:0), PC (16:0/14:0), PC(16:0/18:0), and HHQ for all 403 diseases against the final BOS-grade. Statistics of the linear regression are shown on the plots. 404 b) Box plots of feature abundances of each molecule with final CLAD stages separated by 405 individual diseases (AATD, CF, COPD, IPF). Pearson correlation test (R) is displayed, and *** = 406 p values \leq 0.001.

It is made available under a CC-BY-ND 4.0 International license .

It is made available under a CC-BY-ND 4.0 International license .



408

Figure 4. *Pseudomonas*-derived molecules in subjects that underwent lung transplant due to CF disease. A) Molecular networks of microbial molecules produced by *Pseudomonas* sp. that were identified by GNPS library searching. Each node represents a unique MS/MS spectrum (putative molecule), and connections between nodes are determined by spectral similarity

It is made available under a CC-BY-ND 4.0 International license .

413 (cosine score) from MS/MS alignment. Pie charts represent total feature abundance colored by
414 the underlying disease. B) Percentage of samples that displayed *Pseudomonas*-like molecules
415 in subjects with CF across longitudinal BALF sampling time points and C) dynamics among
416 subjects over time. D) Linear correlation plot displaying the abundance of selected
417 *Pseudomonas* molecules post-LTX and E) its relation to *Pseudomonas spp.* relative abundance
418 in samples.

It is made available under a CC-BY-ND 4.0 International license .

420 **References**

- 421 1 Afonso JE, Júnior, de Campos Werebe E, *et al.* Lung transplantation. *Einstein*422 2015;**13**:297.
- Yusen RD. Technology and Outcomes Assessment in Lung Transplantation. Proceedings
 of the American Thoracic Society. 2009;6:128–36. doi:10.1513/pats.200809-102go
- 425 3 Chambers DC, Cherikh WS, Goldfarb SB, et al. The International Thoracic Organ
- 426 Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-
- 427 fifth adult lung and heart-lung transplant report-2018; Focus theme: Multiorgan
- 428 Transplantation. *J Heart Lung Transplant* 2018;**37**:1169–83.
- 4 Tissot A, Danger R, Claustre J, *et al.* Early Identification of Chronic Lung Allograft
 Dysfunction: The Need of Biomarkers. *Front Immunol* 2019;**10**.
- 431 doi:10.3389/fimmu.2019.01681
- 432 5 DerHovanessian A, Wallace WD, Lynch JP 3rd, *et al.* Chronic Lung Allograft Dysfunction:
 433 Evolving Concepts and Therapies. *Semin Respir Crit Care Med* 2018;**39**:155–71.
- 434 6 Verleden GM, Glanville AR, Lease ED, *et al.* Chronic lung allograft dysfunction: Definition,
 435 diagnostic criteria, and approaches to treatment- A consensus report from the Pulmonary
 436 Council of the ISHLT. *J Heart Lung Transplant* 2019;**38**:493–503.
- Conrad A, Janciauskiene S, Köhnlein T, *et al.* Impact of alpha 1-antitrypsin deficiency and
 prior augmentation therapy on patients' survival after lung transplantation. *Eur Respir J*2017;**50**. doi:10.1183/13993003.00962-2017
- Kneidinger N, Milger K, Janitza S, *et al.* Lung volumes predict survival in patients with
 chronic lung allograft dysfunction. *Eur Respir J* 2017;49. doi:10.1183/13993003.013152016
- Fernandez IE, Heinzelmann K, Verleden S, *et al.* Characteristic patterns in the fibrotic lung.
 Comparing idiopathic pulmonary fibrosis with chronic lung allograft dysfunction. *Ann Am Thorac Soc* 2015;**12 Suppl 1**:S34–41.
- 446 10 Joshua Blatter SS. Lung Transplantation in Cystic Fibrosis: Trends and Controversies.
 447 *Pediatr Allergy Immunol Pulmonol* 2015;**28**:237.
- Walmsley S, Cruickshank-Quinn C, Quinn K, *et al.* A prototypic small molecule database for
 bronchoalveolar lavage-based metabolomics. *Sci Data* 2018;**5**:180060.
- 450 12 Das S, Bernasconi E, Koutsokera A, *et al.* A prevalent and culturable microbiota links
 451 ecological balance to clinical stability of the human lung after transplantation. *Nat Commun*452 2021;**12**:2126.
- 453 13 Combs MP, Wheeler DS, Luth JE, *et al.* Lung microbiota predict chronic rejection in healthy

It is made available under a CC-BY-ND 4.0 International license .

454		lung transplant recipients: a prospective cohort study. Lancet Respir Med 2021;9:601-12.
455	14	Evans CR, Karnovsky A, Kovach MA, et al. Untargeted LC-MS metabolomics of
456		bronchoalveolar lavage fluid differentiates acute respiratory distress syndrome from health.
457		J Proteome Res 2014; 13 :640–9.
458	15	Cribbs SK, Park Y, Guidot DM, et al. Metabolomics of bronchoalveolar lavage differentiate
459		healthy HIV-1-infected subjects from controls. AIDS Res Hum Retroviruses 2014;30:579-
460		85.
461	16	Watzenboeck ML, Gorki A-D, Quattrone F, et al. Multi-omics profiling predicts allograft
462		function after lung transplantation. Eur Respir J 2022;59. doi:10.1183/13993003.03292-
463		2020
464	17	Walter S, Gudowius P, Bosshammer J, et al. Epidemiology of chronic Pseudomonas
465		aeruginosa infections in the airways of lung transplant recipients with cystic fibrosis. Thorax
466		1997; 52 :318–21.
467	18	da Silva RR, Lopes NP, Silva DB. CHAPTER 3:Metabolomics. In: Mass Spectrometry in
468		Chemical Biology. 2017. 57–81.
469	19	Garg N, Kapono C, Lim YW, et al. Mass spectral similarity for untargeted metabolomics
470		data analysis of complex mixtures. Int J Mass Spectrom 2015;377:719–717.
471	20	Quinn RA, Nothias L-F, Vining O, et al. Molecular Networking As a Drug Discovery, Drug
472		Metabolism, and Precision Medicine Strategy. Trends Pharmacol Sci 2017;38:143-54.
473	21	Wang M, Carver JJ, Phelan VV, et al. Sharing and community curation of mass
474		spectrometry data with Global Natural Products Social Molecular Networking. Nat
475		Biotechnol 2016; 34 :828–37.
476	22	Nothias L-F, Petras D, Schmid R, et al. Feature-based molecular networking in the GNPS
477		analysis environment. <i>Nat Methods</i> 2020; 17 :905–8.
478	23	Estenne M, Hertz MI. Bronchiolitis obliterans after human lung transplantation. Am J Respir
479		<i>Crit Care Med</i> 2002; 166 :440–4.
480	24	Pluskal T, Castillo S, Villar-Briones A, et al. MZmine 2: modular framework for processing,
481		visualizing, and analyzing mass spectrometry-based molecular profile data. BMC
482		Bioinformatics 2010;11:395.
483	25	Vázquez-Baeza Y, Pirrung M, Gonzalez A, et al. EMPeror: a tool for visualizing high-
484		throughput microbial community data. Gigascience 2013;2:16.
485	26	Wickham H, Hester J, Chang W. Devtools: Tools to make developing r packages easier. R
486		package version
487	27	Oksanen J, Kindt R, Legendre P, et al. The vegan package. Community ecology package

It is made available under a CC-BY-ND 4.0 International license .

488 2007;**10**:719.

- Lucas SK, Villarreal AR, Ahmad MM, *et al.* Anaerobic Microbiota Derived from the Upper
 Airways Impact Staphylococcus aureus Physiology. *Infect Immun* 2021;**89**:e0015321.
- 29 Callahan BJ, McMurdie PJ, Rosen MJ, *et al.* DADA2: High-resolution sample inference
 492 from Illumina amplicon data. *Nat Methods* 2016;**13**:581–3.
- 30 Raghuvanshi R, Vasco K, Vázquez-Baeza Y, *et al.* High-Resolution Longitudinal Dynamics
 of the Cystic Fibrosis Sputum Microbiome and Metabolome through Antibiotic Therapy. *mSystems* 2020;**5**. doi:10.1128/mSystems.00292-20
- 496 31 Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer 2016.
- 497 32 Oksanen J, Blanchet FG, Kindt R, *et al.* Vegan: community ecology package. R package
 498 version 1.17-4. *http://cran r-project org> Acesso em* 2010;**23**:2010.
- Barr HL, Halliday N, Cámara M, *et al.* Pseudomonas aeruginosa quorum sensing
 molecules correlate with clinical status in cystic fibrosis. *Eur Respir J* 2015;46:1046.
- 34 da Silva RR, Dorrestein PC, Quinn RA. Illuminating the dark matter in metabolomics. Proc.
 502 Natl. Acad. Sci. U. S. A. 2015;**112**:12549–50.
- 503 35 Hoffmann MA, Nothias L-F, Ludwig M, *et al.* High-confidence structural annotation of 504 metabolites absent from spectral libraries. *Nat Biotechnol* 2022;**40**:411–21.
- 36 Jarmusch AK, Wang M, Aceves CM, *et al.* ReDU: a framework to find and reanalyze public
 mass spectrometry data. *Nat Methods* 2020;**17**:901–4.
- 507 37 Ernst M, Kang KB, Caraballo-Rodríguez AM, *et al.* MolNetEnhancer: Enhanced Molecular
 508 Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites* 2019;9.
 509 doi:10.3390/metabo9070144
- 510 38 Berkebile AR, McCray PB, Jr. Effects of airway surface liquid pH on host defense in cystic
 511 fibrosis. *Int J Biochem Cell Biol* 2014;**52**:124.
- S12 39 Nkadi PO, Allen Merritt T, Pillers D-AM. An overview of pulmonary surfactant in the
 neonate: Genetics, metabolism, and the role of surfactant in health and disease. Molecular
 Genetics and Metabolism. 2009;97:95–101. doi:10.1016/j.ymgme.2009.01.015
- 515 40 Chen Z, Zhong M, Luo Y, *et al.* Determination of rheology and surface tension of airway
- 516 surface liquid: a review of clinical relevance and measurement techniques. *Respir Res*517 2019;**20**:1–14.
- 41 Anandan A, Evans GL, Condic-Jurkic K, *et al.* Structure of a lipid A phosphoethanolamine
 transferase suggests how conformational changes govern substrate binding. *Proc Natl Acad Sci U S A* 2017;**114**:2218.
- 521 42 Sohlenkamp C, Geiger O. Bacterial membrane lipids: diversity in structures and pathways.

It is made available under a CC-BY-ND 4.0 International license .

522 *FEMS Microbiol Rev* 2016;**40**:133–59.

523 43 Smith EE, Buckley DG, Wu Z, *et al.* Genetic adaptation by Pseudomonas aeruginosa to the 524 airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 2006;**103**.

525 doi:10.1073/pnas.0602138103

- 526 44 Syed SA, Whelan FJ, Waddell B, *et al.* Reemergence of Lower-Airway Microbiota in Lung
 527 Transplant Patients with Cystic Fibrosis. *Ann Am Thorac Soc* 2016;**13**:2132–42.
- McCort M, MacKenzie E, Pursell K, *et al.* Bacterial infections in lung transplantation. J
 Thorac Dis 2021;**13**:6654–72.
- 530 46 Migiyama Y, Kaneko Y, Yanagihara K, et al. Efficacy of AiiM, an N-acylhomoserine
- 531 lactonase, against Pseudomonas aeruginosa in a mouse model of acute pneumonia.
- 532 Antimicrob Agents Chemother 2013;57. doi:10.1128/AAC.00456-13





a)

medRxiv preprint doi: https://doi.org/10.1101/2022.11.16.22281980; this version posted November 18, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.



CLAD stages

Diseases



Quinolones nomenclature

HQNO: 2-heptyl-4-quinolone-N-oxide HNQ: 2-nonyl-4-quinolone HHQ: 4-hydroxy-2-heptylquinolone Color information in sub-networks

Pie chart colors represents feature abundance based on diseases
AATD CF COPD IPF
Annotated by GNPS: CEdge width represent spectral similarity between nodes (cosine score)
0.70 - 1.00

