

1 **BILE ACIDS IN LOWER AIRWAYS AS A NOVEL INDICATOR OF AIRWAY**
2 **MICROBIOTA CHANGES IN CHRONIC OBSTRUCTIVE PULMONARY**
3 **DISEASE.**

4

5 Jose A. Caparrós-Martín^{1,2,§}, Montserrat Saladié^{2,§}, S. Patricia Agudelo-
6 Romero^{1,3}, Kristy S. Nichol⁴, F. Jerry Reen^{5,6}, Yuben Moodley^{7,8,9,10}, Siobhain
7 Mulrennan^{10,11}, Stephen M. Stick^{1,3,12}, Peter A Wark⁴ and Fergal O’Gara^{1,2,13}.

8

9 ¹Wal-yan Respiratory Research Centre. Telethon Kids Institute, Perth, Western
10 Australia, Australia.

11 ²Curtin Health Innovation Research Institute (CHIRI), Curtin University, Perth,
12 Western Australia, Australia.

13 ³The University of Western Australia, Perth, Western Australia, Australia.

14 ⁴Immune Health Program, Hunter Medical Research Institute, University of
15 Newcastle, Newcastle, New South Wales, Australia.

16 ⁵School of Microbiology, University College Cork, Cork, Ireland.

17 ⁶Synthesis and Solid-State Pharmaceutical Centre, University College Cork,
18 Cork, Ireland.

19 ⁷Centre for Respiratory Health, School of Biomedical Science, The University of
20 Western Australia, Nedlands, Western Australia, Australia.

21 ⁸Cell Biology Group, Institute for Respiratory Health, Nedlands, Western
22 Australia, Australia.

23 ⁹Department of Respiratory Medicine, Fiona Stanley Hospital, Murdoch,
24 Western Australia, Australia.

25 ¹⁰UWA Medical School, The University of Western Australia, Perth, Western
26 Australia, Australia.

27 ¹¹Cystic Fibrosis Research Unit, Institute for Respiratory Health, Nedlands,
28 Western Australia, Australia.

29 ¹²Department of Respiratory Medicine, Princess Margaret Hospital for Children,
30 Perth, Western Australia, Australia.

31 ¹³BIOMERIT Research Centre, School of Microbiology, University College Cork,
32 Cork, Ireland.

33

34 [§] These authors contributed equally to the manuscript

35

36 Correspondence to:

37 Professor Fergal O’Gara; BIOMERIT Research Centre, School of Microbiology,

38 University College Cork, T12 K8AF Cork, Ireland. Tel.: +353-21-4901315; Fax:

39 +353-21-4275934 Email: f.ogara@ucc.ie

40

41 **Abstract**

42 **Background.** Chronic obstructive pulmonary disease (COPD) is a complex
43 disorder with a high degree of interindividual variability. Gastrointestinal
44 dysfunction is common in COPD patients and has been proposed to influence
45 the clinical progression of the disease. Using the presence of bile acid(s) (BA) in
46 bronchoalveolar lavage fluid (BAL) as a marker of gastric aspiration, we
47 evaluated the relationships between BAs, clinical outcomes, and bacterial lung
48 colonisation.

49 **Methods.** We used BAL specimens from a cohort of COPD patients and
50 healthy controls. Bile acids were profiled and quantified in BAL supernatants
51 using mass spectrometry. Microbial DNA was extracted from BAL cell pellets
52 and quantified using qPCR. We profiled the BAL microbiota using an amplicon
53 sequencing approach targeting the V3-V4 region of the 16S rRNA gene.

54 **Results.** Detection of BAs in BAL was more likely at earliest clinical stages of
55 COPD and was independent of the degree of airway obstruction. BAL
56 specimens with BAs demonstrated higher bacterial biomass and lower diversity.
57 Likewise, the odds of recovering bacterial cultures from BAL were higher if BAs
58 were also detected. Detection of BAs in BAL was not associated with either
59 inflammatory markers or clinical outcomes. We also observed different bacterial
60 community types in BAL, which were associated with different clinical groups,
61 levels of inflammatory markers, and the degree of airway obstruction.

62 **Conclusion.** Detection of BAs in BAL was associated with different parameters
63 of airway ecology. Further studies are needed to evaluate whether BAs in BAL
64 can be used to stratify patients and for predicting disease progression
65 trajectories.

66

67 **Keywords.** Microbiota, chronic obstructive pulmonary disease, Bile acids,
68 gastroesophageal reflux disease.

69

70 **Introduction**

71 Chronic obstructive pulmonary disease (COPD) is a progressive lung condition
72 associated with chronic inflammation, airflow obstruction, and irreversible
73 damage of the alveolar walls^{1,2}. The development of COPD is linked to long-
74 term exposure to tobacco smoke and airborne pollutants that irritate the
75 airways². The clinical evolution of COPD is dictated by worsening episodes of
76 variable aetiology and severity called exacerbations². Other factors, such as
77 microbial dynamics and airway colonisation, airway hyper-reactivity, and
78 immune and gastroesophageal dysfunction have all been shown to influence
79 disease severity and progression².

80 The prevalence of functional gastrointestinal pathology in patients with chronic
81 lung disease is high³. A common extrapulmonary manifestation in patients with
82 COPD is gastroesophageal reflux disease (GERD)⁴. Numerous epidemiological
83 studies have shown a robust relationship between GERD and increased
84 incidence of COPD exacerbations⁵⁻⁸. This suggests that reflux may modulate
85 disease severity in COPD. The mechanisms through which GERD could
86 aggravate the pathophysiology of COPD are unknown, but it may involve the
87 translocation of gastric metabolites and/or microorganisms into the lower
88 airways through a reflux-microaspiration process⁹.

89 Bile acids (BAs) are host metabolites involved in fat digestion. Apart from their
90 role as fat emulsifiers, BAs have antimicrobial, endocrine and immune
91 regulatory properties, and modulate the expression of pathogenesis-related
92 genes in bacterial pathogens^{10,11}. Several studies have identified BAs in
93 bronchoalveolar lavage fluid (BAL) obtained from patients with different
94 respiratory pathologies³. When detected in BAL, BAs have been associated
95 with pathophysiologic events in chronic lung disease. Thus, the presence of
96 BAs in BAL has been linked to airway inflammation in patients with cystic
97 fibrosis¹²⁻¹⁵, lung transplant (LTx) recipients¹⁶⁻¹⁸, and children with respiratory
98 symptoms¹⁶. In LTx recipients, detection of BAs in BAL is also a risk factor for
99 allograft dysfunction and mortality¹⁷.

100 The lung microbiota originates from the aspiration of pharyngeal commensals¹⁹,
101 through a process that also calibrates the lung immune tone²⁰. In COPD, the
102 composition of the lung microbiota represents a source of clinical heterogeneity.
103 Different microbial assemblages or endotypes are linked to specific COPD
104 clinical phenotypes, which determine treatment response and disease
105 progression trajectories^{21,22}. Whether and how bacterial aspiration is clinically
106 relevant for the progression of COPD is unknown.

107 In this study, we evaluated associations between detection of BAs in BAL,
108 parameters of airway ecology and clinical outcomes. We used a cohort of BAL
109 specimens collected from clinically stable COPD patients at different stages of
110 the disease, and healthy controls. Detection of BAs in BAL was robustly
111 associated with parameters of airway ecology. Overall, our data suggest that
112 detection of BAs in BAL indicates processes associated with airway microbial
113 colonisation in COPD.

114

115 **Results**

116 **Study population**

117 This is a retrospective study of 166 bronchoalveolar lavage (BAL) specimens
118 collected from 108 patients with COPD (GOLD1-4) and healthy non-smokers
119 (nSC, 38) and healthy smoker (SC, 20) controls. Bronchoalveolar lavage
120 samples were collected as previously described²³. Patients with COPD were
121 stable, with no change in medications or acute exacerbations in the previous 6
122 weeks. Healthy controls had spirometry in the normal range (FEV1>80%
123 predicted, FEV1/FVC > 0.7), and no history of chronic lung disease. Patient
124 demographics are provided in Table 1.

125

126 **Detection of bile acids in BAL is associated with early stages of COPD.**

127 We detected bile acids (BAs) in 67 (40.4%) BAL specimens (median
128 concentration [interquartile range (IQR)], 0.24 [0.13-0.47] μM) (Table S1). The
129 concentration of BAs in BAL from COPD patients was higher than in healthy
130 donors (median [IQR] in controls, 0 [0-0] μM ; median [IQR] in COPD patients,
131 0.007 [0-0.09] μM , Wilcoxon Rank Sum Test (WRST), $p=0.0006$, $r=0.27$)
132 (Figure S1). In our cohort, detection of BAs in BAL was associated with patient
133 age but not with the degree of airway obstruction (Figure S2).

134 Detection of BAs was more likely in BAL collected from COPD patients than
135 from healthy controls (odds ratio (OR) [95% confidence interval], 3.98 [1.90-
136 8.33] $p=0.0002$). Eight of 38 non-smoker healthy controls (21%), 4 of 20 smoker
137 healthy controls (20%), 14 of 25 COPD patients from GOLD stage I (56%), 19
138 of 32 COPD patients from GOLD stage II (59%), 18 of 36 COPD patients from
139 GOLD stage III (50%), and 4 of 15 COPD patients from GOLD stage IV (27%),
140 had BAs in BAL. Compared to the nSC group, the likelihood for detecting BAs
141 in BAL was increased in early stages of COPD (OR [95% confidence interval],
142 GOLD1 4.77[1.57-14.48] $p=0.006$, GOLD2 5.48[1.91-15.69] $p=0.001$, GOLD3
143 3.75[1.36-10.37] $p=0.01$, GOLD4 1.36[0.34-5.45] $p=0.66$) (Figure 1). We
144 obtained similar results after controlling for the degree of airway obstruction
145 (FEV_1/FVC ratio) or patient age (Figure S3).

146

147 **Relationships between the presence of BAs in BAL and clinical outcomes.**

148 COPD patients with or without BAs in BAL, demonstrated similar levels of white
149 cells in their lungs (median ($\times 10^6$ cells/mL) [IQR], BAs detected 0.36 [0.25-0.57],
150 no BAs detected 0.18 [0.16-0.63], WRST, $p=0.3$, $r=0.24$), with no differences in
151 viability (median ($\times 10^6$ cells/mL) [median (% of viable cells) [IQR]], BAs
152 detected 90 [83-93], no BAs detected 92 [77-95], WRST, $p=0.75$, $r=0.08$).
153 Similarly, there was no difference in the proportion of neutrophils in BAL
154 (median (%) [IQR], BAs detected 68.4 [56.3-84.3], no BAs detected 62.6 [43.3-
155 74.3], WRST, $p=0.1$, $r=0.16$). Conversely, detection of BAs was linked to a
156 higher bacterial burden in BAL (median (Log_{10} 16S copies μL^{-1} of DNA extract)
157 [IQR], BAs detected 4.54 [3.98-5.30], no BAs detected 4.14 [3.60-4.76], WRST,
158 $p=0.01$, $r=0.25$) (Figure S4). In the case of the healthy control groups (smokers
159 and non-smokers), detection of BAs in BAL was not linked to differences in the
160 proportion of infiltrated neutrophils, number and viability of white blood cells, or
161 bacterial load in BAL (Figure 2A).

162 In patients with COPD, the presence of BAs in BAL was not correlated with
163 clinical outcomes, including the ratio of FEV_1 to FVC (median [IQR], BAs
164 detected 57 [48-67], no BAs detected 52 [42-64], WRST, $p=0.32$, $r=0.1$), quality
165 of life (COPD assessment test, CAT) (median [IQR], BAs detected 13 [10-16],
166 no BAs detected 15 [7-19], WRST, $p=0.95$, $r=0.01$), and either number of
167 exacerbations (median [IQR], BAs detected 4 [2-7], no BAs detected 3 [1-

168 5],WRST, $p=0.45$, $r=0.11$), number of antibiotic rounds (median [IQR], BAs
169 detected 2 [1-5], no BAs detected 2 [1-4], WRST, $p=0.58$, $r=0.08$), or number of
170 oral corticosteroids rounds (median [IQR], BAs detected 1 [0-2], no BAs
171 detected 0 [0-1],WRST, $p=0.44$, $r=0.12$) (Figure S5).

172

173 **Associations between the COPD BAL microbiota and clinical outcomes.**

174 We analysed the bacterial profiles of 104 BAL samples with bacterial biomass
175 above the cut-off defined by the negative extraction controls (Figure S6). These
176 taxonomic profiles were dominated by operational taxonomic units (OTUs)
177 assigned to the *Streptococcus* (mean relative abundance (standard deviation),
178 20.19% (17.79)) and *Haemophilus* (13.22% (28.58)) taxa (Figure S7).

179 In our cohort, alpha (within sample) diversity was negatively associated with
180 bacterial biomass (β -0.44, Adjusted R^2 0.14, $F(1,102)=17.20$, $p=0.00007$) and
181 the percentage of neutrophils in BAL (β -0.01, Adjusted R^2 0.10, $F(1,97)=12.41$,
182 $p=0.0006$) (Figure S8). We obtained similar results when the data from healthy
183 controls (both smokers and non-smokers) was excluded from the linear models
184 (Figure S9). Despite the observed relationships between microbial parameters
185 and inflammation in BAL from COPD patients, the degree of airway obstruction
186 was not predicted by either the bacterial diversity (Shannon index, β 2.66,
187 Adjusted R^2 0.02, $F(1,70)=2.17$, $p=0.14$), or the bacterial biomass in BAL
188 (Log₁₀ 16S copies μL^{-1} of DNA extract, β -1.70, Adjusted R^2 0.00,
189 $F(1,70)=0.71$, $p=0.40$) (Figure S10A-B). Likewise, quality of life or number of
190 exacerbations were not correlated with either bacterial load or diversity in BAL
191 (Figure S10C-F). Conversely, BAL specimens containing BAs demonstrated
192 reduced diversity (median [IQR], BAs detected 1.90 [0.78-2.50], no BAs
193 detected 2.42 [1.82-2.77]; WRST, $p=0.01$, $r=0.24$), and increased odds for
194 isolating microorganisms in cultures from BAL (any microorganism, OR
195 2.41[1.15-5.02] $p=0.02$) (Figure 2B-C).

196 We used Dirichlet Multinomial Mixtures to classify the BAL-associated microbial
197 communities and observed 3 different clusters or ecotypes (Figure 3 and Figure
198 S11). These 3 ecotypes showed contrasting levels of BAL-associated bacterial
199 biomass (Figure 3C). At community level, two OTUs assigned to the
200 *Streptococcus* and *Haemophilus* taxa were the more abundant entities defining

201 each community-based cluster (Figure S12). Interestingly, each ecotype was
202 associated with different clinical outcomes as well as specific clinical groups
203 (Figure 3 and Figure S13). Ecotypes I and II occurred in samples from both
204 healthy controls and COPD patients, although it is true that ecotype II was less
205 likely in healthy controls (Figure S13). Both ecotype I and II were dominated by
206 *Streptococcus*, however ecotype II had a less evenly distributed community
207 composition (Figure 3B). Clinically, the transition from ecotype I to ecotype II
208 was associated with a progressive increase in airway inflammation and higher
209 degree of obstructive impairment on spirometry (Figure 3D-E). On the other
210 hand, the *Haemophilus*-dominated community (ecotype III) was only observed
211 in BAL specimens collected from COPD patients (Figure S13). Ecotype III was
212 also associated with a higher percentage of neutrophils in BAL (Figure 3E).
213 None of the 3 ecotypes was significantly associated with either specific GOLD
214 stages or BA detection in BAL (Figure S13).

215

216 **Discussion**

217 Gastrointestinal manifestations are commonly observed in patients with COPD,
218 and they have been linked to disease heterogeneity⁵. Thus, gastroesophageal
219 reflux disease (GERD), which could lead to the aspiration of gastric contents, is
220 associated with a “frequent exacerbator” phenotype⁵⁻⁸. In patients with COPD,
221 several of the physiological barriers preventing the gastric contents reaching the
222 airways are impaired³. Moreover, most COPD patients smoke or are frequently
223 treated with beta-agonists and theophyllines, which affect the physiology of the
224 oesophageal sphincters³. These functional defects are therefore expected to
225 facilitate reflux and aspiration of gastric contents, and they likely represent a
226 functional link between the gastrointestinal tract and respiratory disease. In this
227 study, we evaluated the clinical significance of the presence of BAs in BAL and
228 their impact on BAL-associated microbiota. Bile acids are gastric metabolites
229 that, when observed in the lower airways, are considered a gastric marker for
230 aspiration⁹. In our cohort, the likelihood of detecting BAs in BAL was
231 significantly increased in patients with early-stage COPD disease compared to
232 healthy individuals. This observation was independent of the degree of airway
233 obstruction present but was positively linked to patient age. Past publications
234 have reported contrasting results regarding a relationship between the

235 presence of BAs in BAL, and lung function^{12,14,16,24}. This discrepancy between
236 studies suggests that while lung deterioration could facilitate BAs detection in
237 BAL (e.g. mucociliary clearance impairment), extrapulmonary mechanisms are
238 also involved³. We also detected BAs in 20-21% of BAL samples collected from
239 healthy controls, and this was independent of smoking status. A previous study
240 reported BAs in 13% of sputum samples from healthy individuals¹³, supporting
241 the notion that BAs reach the airways in the absence of a respiratory pathology.
242 Likewise, this 10-20% detection range in healthy subjects could be explained by
243 the estimated 10-30% prevalence of GERD in the Western population²⁵.
244 Persistent airway inflammation is a common feature in chronic lung disease.
245 Although tobacco smoke is the major driver of inflammation in COPD lungs,
246 inflammatory markers remain elevated even after the cessation of smoking²⁶.
247 Thus, additional factors are likely involved in maintaining the inflammatory
248 milieu in COPD lungs. Bile acids have been robustly linked to inflammation and
249 poor clinical outcomes in different respiratory disorders^{9,12-14,16-18,24}, and they
250 could provide a link between GERD and COPD disease severity⁵. In our study,
251 the percentage of neutrophils in COPD BAL regardless of BAs presence was
252 comparable, suggesting that BAs are not involved in establishing the
253 inflammatory landscape of COPD lungs. Similarly, we did not observe a
254 relationship between the presence of BAs in BAL and clinical outcomes. These
255 results are in line with a previous study whose findings did not support a role of
256 BAs in COPD exacerbations²⁷. We however observed a robust association
257 between BAs and ecological parameters in BAL. Aspiration can shape the
258 COPD lung microbiota²⁸, and is a mechanism for pulmonary colonisation¹⁹.
259 Likewise, bacterial, and viral infections are known triggers of exacerbations in
260 COPD²⁹. In our study, COPD BAL samples with BAs demonstrated higher
261 bacterial load and lower diversity than specimens without BAs. Additionally, the
262 probability for a positive microbial culture from BAL was significantly increased if
263 BAs were also detected. Considering this, future studies are required to clarify
264 whether the known relationship between GERD and COPD exacerbations could
265 be explained by reflux increasing the risk of aspiration of microbial pathogens.
266 The lung microbiome plays an important role in the progression of COPD^{21,22}.
267 We found that BAL bacterial biomass and diversity were negatively correlated
268 with the percentage of infiltrated neutrophils. However, clinical outcomes were

269 independent of bacterial community descriptors in BAL. This unexpected
270 observation supports previous studies demonstrating that clinical responses in
271 COPD depends on particular microbial community signatures^{21,22}, rather than
272 on the overall structure of the bacterial community.

273 We also examined the microbial communities present in BAL. Using cluster
274 analysis and likelihood estimators, we identified three community types, that
275 were correlated with contrasting degrees of airway obstruction and
276 inflammation, and different clinical groups. We found that *Haemophilus*-
277 dominated communities were characterized by elevated biomass, and were
278 also linked to the highest percentage of neutrophils in BAL. This community
279 type was exclusively observed in COPD BAL samples. This result is in line with
280 previous work in COPD demonstrating that *Haemophilus* dominance is
281 consistently associated with neutrophilic inflammation^{21,22}. We also identified
282 two *Streptococcus*-dominated microbial community types, which could
283 represent different states of the balanced microbiome profile previously
284 described in COPD sputum^{21,22}. *Streptococcus* dominance was observed in
285 BAL from both healthy controls and COPD patients, and it was not linked to
286 specific clinical stages of the disease. Interestingly, both ecotypes showed
287 contrasting patterns of microbial distribution, with the less evenly distributed
288 signature associated with a progressive increase in airway obstruction and
289 inflammation. Detection of BAs in BAL did not correlate with specific microbial
290 assemblages. This result suggests that although bacteria and BAs could access
291 the airways through the same physiological process, additional factors
292 determine the type of bacterial community thriving in the lungs.

293 An obvious limitation of our study is its observational nature, which does not
294 allow causality to be established. Likewise, participants did not undergo
295 gastroesophageal reflux monitoring. Thus, we cannot confirm the physiological
296 alteration mediating BAs intrusion into the airways. Although the physiological
297 process underlying the translocation of BAs into the lungs is not relevant for the
298 associations shown in the present study, BAs in BAL are considered a marker
299 of gastric aspiration⁹. Lastly, BAs and the microbial profiles could be
300 confounded by salivary contamination during bronchoscopy. Although we
301 cannot rule out this possibility, our results demonstrating an association

302 between detection of BAs and specific clinical groups are unlikely to be driven
303 by a contamination signal.

304 Overall, in this cross-sectional study, we show that detection of BAs in BAL is
305 associated with early stages of COPD, and correlated with parameters of airway
306 ecology, but not with clinical outcomes. Future longitudinal studies are needed
307 to determine whether the presence of BAs in BAL at early clinical stages of
308 COPD is a useful approach to stratify COPD patients (e.g. to capture those
309 subjects in a worse disease progression trajectory).

310

311 **FIGURE LEGENDS.**

312 **Figure 1.** Forest plots showing the odds of detecting BAs in BAL in the different
313 population groups. Odds ratio are represented with 95% pointwise confidence
314 interval.

315

316 **Figure 2. A-B.** Boxplots overlaid with density plots (blue) showing the
317 relationship between BAs detection in BAL and bacterial load (**A**), and the
318 diversity of the BAL-associated microbial communities (**B**). Individual values for
319 each BAL specimen are represented with red dots. Notches in the boxplots
320 represent 95% confidence interval for the sample median. Statistical
321 significance was assessed using the Wilcoxon rank sum test. For each
322 comparison, the corresponding p-value (p) is shown in the graph. **C.** Marginal
323 effect of BAs detection in BAL on the odds of observing positive microbial
324 cultures from BAL with pointwise 95% confidence intervals. Predicted
325 probabilities were calculated from a logistic regression model.

326

327 **Figure 3. A.** Scatter plot showing the projection of the BAL-associated bacterial
328 compositional profiles in the first two components of the principal component
329 analysis. Samples are labelled accordingly with the cluster membership. **B.**
330 Heatmap summarising the top 10 OTUs ordered based on their contribution to
331 the Dirichlet components. Narrow columns represent the square-root of the
332 OTU relative abundance, while wide columns represent the mean value of the
333 Dirichlet component for each mixture. **C-E.** Barplots overlaid with density curves
334 representing differences in bacterial load (**C**), FEV1/FVC ratio (**D**), and the
335 percentage of neutrophils in BAL (**E**), relative to the indicated microbial

336 community cluster. Red dots represent observed values for each BAL
337 specimen. Statistical significance was assessed using the Wilcoxon rank sum
338 test and p-values corrected for multiple comparisons using the Bonferroni
339 method.

340

341 **Methods**

342 **Isolation of microbial DNA from BALF, 16S metabarcoding, sequencing of** 343 **the amplicon pools, and processing of the sequencing data.**

344 We obtained cell pellets from 163 out of the 166 BALF specimens through high-
345 speed centrifugation of 1-2 mL of BALF. Reagent-related contaminants were
346 controlled by sequencing negative extraction controls. Extraction of microbial
347 DNA and amplicon-based bacterial profiling targeting the V3-V4 region of the
348 16S rRNA gene were performed blinded to clinical data as we described³⁰. In
349 this project, we obtained 19,236,207 single reads with a good base calling
350 accuracy (96% of sequences with mean Phred-like Q-score greater than or
351 equal to 30). Raw sequencing data was processed as we described³⁰. Quality-
352 based pre-processing resulted in 8,538,607 paired-end reads (average length
353 405 bp), with >94% of the reads having a mean Phred-like Q-score greater than
354 or equal to 35. Pre-processed joint reads were analysed following the SILVAngs
355 pipeline^{31,32} as we described³⁰. For taxonomic classification we used the last
356 release of the standardised SILVA SSU taxonomy (release 138.1) as
357 reported³⁰. Following this processing pipeline, amplicon sequencing data was
358 assigned to 1,280 Operational Taxonomic Units (OTUs).

359

360 **Removal of the potential background contaminants determined by** 361 **negative extraction controls.**

362 Bacterial burden in negative processing controls was used to determine the
363 background noise associated with our experimental conditions. For 17 DNA
364 extracts, we did not have enough DNA for assessing bacterial load, and the
365 taxonomy profiles from these BALF samples were therefore removed from the
366 final dataset. We set the background cut-off to 4,100 16S DNA copies μL^{-1} of
367 DNA extract, which lies three standard deviations from the mean bacterial load
368 in the negative extraction controls (mean 1,722.58, standard deviation 777.86,
369 16S DNA copies μL^{-1} DNA extract) (Figure S14). In our dataset, 104 DNA

370 extracts from BAL had bacterial burden above the background cut-off. These
371 104 DNA extracts were therefore considered for further analysis on the basis
372 that, the associated microbial profiles, are more likely to represent a true
373 biological signal. The taxonomic table was then processed to eliminate
374 singletons, reads classified as Chloroplast (mean relative abundance (standard
375 deviation), 0.005% (0.03)), Mitochondria (0.00004% (0.0002)) or Eukaryota taxa
376 (0.0000006% (0.000009)), and no classified reads (0.001% (0.002)). Taxa
377 representing less than 0.01% across all samples were also removed. Potential
378 contaminants introduced during the processing of the samples were estimated
379 using the R package *decontam*³³ as we previously described³⁰. We detected
380 and removed from the OTU table, 8 OTUs potentially arising from
381 environmental contamination (Table S2). These filtering steps removed 1,127
382 OTUs from the unfiltered dataset, resulting in a OTU table with 153 unique
383 taxonomy paths. Procrustes analysis confirmed that these filtering steps did not
384 significantly impact the overall structure of the unfiltered microbial profiles
385 (Table S3 and Figure S16). We obtained a good correspondence between
386 clinical microbiology data from BAL cultures and the taxonomic profiles obtained
387 using molecular methods (Table S4).

388

389 **Quantitative analysis of bacterial burden in BAL.**

390 We used a TaqMan® probe to estimate bacteria burden in DNA extracts as
391 described in our previous publication³⁰.

392

393 **Quantitative profiling of bile acids in BAL.**

394 We extracted bile acids from spun BAL supernatants using solid-phase
395 separation¹². For identification and quantification of the different bile acid
396 analytes we used mass spectrometry as we previously described^{12,34}.

397

398 **Statistical analyses.**

399 Statistical analyses and graphical representations were done in R (Version
400 4.0.2). Normality of data distribution the data was evaluated using the Shapiro-
401 Wilk test. When data did not follow a normal distribution, we evaluated
402 differences between group medians using the Wilcoxon Rank Sum Test. Linear
403 models were computed in R using the *lm* function. The assumption of

404 homoscedasticity of the residuals and the absence of high influential data points
405 were confirmed using regression diagnostics. Shannon diversity index was
406 calculated with the *diversity* function of the R package *vegan*. For cluster
407 analysis of the BAL-associated bacterial communities we used Dirichlet
408 Multinomial Mixtures models³⁵. The probability threshold for statistical
409 significance was set at 0.05.

410

411 **Data availability.**

412 All the sequencing data generated in this study have been deposited in the
413 Sequencing Read Archive under the Bioproject accession PRJNA820972.

414

415 **Ethics statement.**

416 Ethical approval was obtained from the Hunter New England LHD Ethics
417 committee [Mechanisms of Inflammatory airways disease (05/08/10/3.09)].

418

419 **Acknowledgments.**

420 We gratefully acknowledge the patients and participants for donating BAL fluid
421 to support this study. This work was funded in part by grants awarded by the
422 Health Research Board (HRB-ILP-POR-2019-004), the Irish Thoracic Society
423 (MRCG-2018-16 and MRCG-2014-6), the US Cystic Fibrosis Foundation
424 (OGARA1710), Australian NHMRC 2020 Synergy grant (APP1183640), the
425 Glenn Brown Memorial Grant 2017 (The Institute for Respiratory Health, Perth,
426 Australia), the European Commission (EU-634486), Science Foundation Ireland
427 (SSPC-2/PharM5, 13/TIDA/B2625, 14/TIDA/2438 and 15/TIDA/2977) and
428 Enterprise Ireland Commercialisation Fund (CF-2017-0757-P). J.A.C.-M. thanks
429 the director of the Curtin Health and Innovation Research Institute (CHIRI)
430 Professor John Mamo and the CHIRI laboratory manager Dr Rob Stuart for
431 granting access to the CHIRI laboratory facilities, and the mass-spectrometry
432 officer Dr Ben Hunter for providing technical support with the mass
433 spectrometer.

434

435 **Author Contributions:** Conceptualization, F.O.G., J.A.C.-M and F.J.R.;
436 Methodology, M.S. and J.A.C.-M.; Formal Analysis, J.A.C.-M. and S.P.A.-R.;
437 Investigation, J.A.C.-M. and M.S.; Resources, P.A.W., Y.M., S.M. and S.M.S.;

438 Writing-Original Draft Preparation, J.A.C.-M. and F.O.G.; Writing – Review &
439 Editing, J.A.C.-M., M.S., S.P.A.-R., N.K., F.J.R., Y.M., S.M., S.M.S., P.A.W. and
440 F.O.G.; Visualization, J.A.C.-M.

441

442 **Conflicts of Interests:** The authors declare no conflict of interest, financial or
443 otherwise. The research funders had no role in study design, data collection
444 and analysis, decision to publish, or preparation of the manuscript.

445

446

	FEV1/FVC, mean (SD); number of patients	Age, mean (SD); number of patients	Sex (% males); number of patients
GOLD1	72.29(7.55); 21	64.52(11.52); 25	40%; 25
GOLD2	57.77(10.70); 32	72.06(7.14); 31	53%; 32
GOLD3	49.53(11.96); 36	69.03(10.74); 36	67%; 36
GOLD4	34.71(9.31); 14	61.93(10.85); 15	60%; 15
nSC	76.15(7.64); 33	57.63(17.12); 38	42%; 38
SC	75.76(13.61); 17	51.85(18.22); 20	55%; 20
Statistical test	$F=53.91$, ***	$F=9.17$, **	$\chi^2 10.82$, n.s.

447

448 **Table 1. Patient demographics.** We used a Joint F-test for assessing
449 differences in the degree of airway obstruction and patient age. A chi-square
450 test of independence was used to evaluate differences in patient sex across the
451 different clinical groups. **, $p<0.01$; ***, $p<0.001$.

452

453 **References:**

454

455

- 456 1. Christenson SA, Smith BM, Bafadhel M, Putcha N. Chronic obstructive
457 pulmonary disease. *Lancet* 2022; **399**(10342): 2227-42.
- 458 2. GOLD. Global Strategy for the Diagnosis, Management and Prevention of
459 COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2021.; 2021.
- 460 3. Houghton LA, Lee AS, Badri H, DeVault KR, Smith JA. Respiratory disease
461 and the oesophagus: reflux, reflexes and microaspiration. *Nat Rev Gastroenterol*
462 *Hepatol* 2016; **13**(8): 445-60.
- 463 4. Casanova C, Baudet JS, del Valle Velasco M, et al. Increased gastro-
464 oesophageal reflux disease in patients with severe COPD. *Eur Respir J* 2004; **23**(6):
465 841-5.
- 466 5. Huang C, Liu Y, Shi G. A systematic review with meta-analysis of
467 gastroesophageal reflux disease and exacerbations of chronic obstructive pulmonary
468 disease. *BMC Pulm Med* 2020; **20**(1): 2.
- 469 6. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to exacerbation in chronic
470 obstructive pulmonary disease. *N Engl J Med* 2010; **363**(12): 1128-38.
- 471 7. Terada K, Muro S, Sato S, et al. Impact of gastro-oesophageal reflux disease
472 symptoms on COPD exacerbation. *Thorax* 2008; **63**(11): 951-5.
- 473 8. Rascon-Aguilar IE, Pamer M, Wludyka P, et al. Role of gastroesophageal reflux
474 symptoms in exacerbations of COPD. *Chest* 2006; **130**(4): 1096-101.
- 475 9. Lee AS, Lee JS, He Z, Ryu JH. Reflux-Aspiration in Chronic Lung Disease. *Ann*
476 *Am Thorac Soc* 2020; **17**(2): 155-64.

- 477 10. Reen FJ, Flynn S, Woods DF, et al. Bile signalling promotes chronic respiratory
478 infections and antibiotic tolerance. *Sci Rep* 2016; **6**: 29768.
- 479 11. Chen ML, Takeda K, Sundrud MS. Emerging roles of bile acids in mucosal
480 immunity and inflammation. *Mucosal Immunol* 2019; **12**(4): 851-61.
- 481 12. Caparrós-Martín JA, Saladié M, Agudelo-Romero P, et al. Detection of Bile
482 Acids in Bronchoalveolar Lavage Fluid Defines Early Pathobiological Events in Infants
483 with Cystic Fibrosis. 2022. s. Available at
484 SSRN: <https://ssrn.com/abstract=4081064> or <http://dx.doi.org/10.2139/ssrn.4081064>
485
- 486 13. Pauwels A, Decraene A, Blondeau K, et al. Bile acids in sputum and increased
487 airway inflammation in patients with cystic fibrosis. *Chest* 2012; **141**(6): 1568-74.
- 488 14. Caparros-Martin JA, Flynn S, Reen FJ, et al. The Detection of Bile Acids in the
489 Lungs of Paediatric Cystic Fibrosis Patients Is Associated with Altered Inflammatory
490 Patterns. *Diagnostics (Basel)* 2020; **10**(5).
- 491 15. Flynn S, Reen FJ, Caparros-Martin JA, et al. Bile Acid Signal Molecules
492 Associate Temporally with Respiratory Inflammation and Microbiome Signatures in
493 Clinically Stable Cystic Fibrosis Patients. *Microorganisms* 2020; **8**(11).
- 494 16. Rosen R, Lurie M, Kane M, et al. Risk Factors for Bile Aspiration and its Impact
495 on Clinical Outcomes. *Clin Transl Gastroenterol* 2021; **12**(12): e00434.
- 496 17. Urso A, Leiva-Juarez MM, Briganti DF, et al. Aspiration of conjugated bile
497 acids predicts adverse lung transplant outcomes and correlates with airway lipid and
498 cytokine dysregulation. *J Heart Lung Transplant* 2021; **40**(9): 998-1008.
- 499 18. Vos R, Blondeau K, Vanaudenaerde BM, et al. Airway colonization and gastric
500 aspiration after lung transplantation: do birds of a feather flock together? *J Heart Lung*
501 *Transplant* 2008; **27**(8): 843-9.
- 502 19. Dickson RP, Erb-Downward JR, Freeman CM, et al. Bacterial Topography of
503 the Healthy Human Lower Respiratory Tract. *mBio* 2017; **8**(1).
- 504 20. Wu BG, Sulaiman I, Tsay JJ, et al. Episodic Aspiration with Oral Commensals
505 Induces a MyD88-dependent, Pulmonary T-Helper Cell Type 17 Response that
506 Mitigates Susceptibility to *Streptococcus pneumoniae*. *Am J Respir Crit Care Med*
507 2021; **203**(9): 1099-111.
- 508 21. Wang Z, Locantore N, Haldar K, et al. Inflammatory Endotype-associated
509 Airway Microbiome in Chronic Obstructive Pulmonary Disease Clinical Stability and
510 Exacerbations: A Multicohort Longitudinal Analysis. *Am J Respir Crit Care Med* 2021;
511 **203**(12): 1488-502.
- 512 22. Dicker AJ, Huang JTJ, Lonergan M, et al. The sputum microbiome, airway
513 inflammation, and mortality in chronic obstructive pulmonary disease. *J Allergy Clin*
514 *Immunol* 2021; **147**(1): 158-67.
- 515 23. Lokwani R, Wark PAB, Baines KJ, Barker D, Simpson JL. Hypersegmented
516 airway neutrophils and its association with reduced lung function in adults with
517 obstructive airway disease: an exploratory study. *BMJ Open* 2019; **9**(1): e024330.
- 518 24. Blondeau K, Mertens V, Vanaudenaerde BA, et al. Gastro-oesophageal reflux
519 and gastric aspiration in lung transplant patients with or without chronic rejection. *Eur*
520 *Respir J* 2008; **31**(4): 707-13.
- 521 25. El-Serag HB, Sweet S, Winchester CC, Dent J. Update on the epidemiology of
522 gastro-oesophageal reflux disease: a systematic review. *Gut* 2014; **63**(6): 871-80.
- 523 26. Willemsse BW, ten Hacken NH, Rutgers B, Lesman-Leegte IG, Postma DS,
524 Timens W. Effect of 1-year smoking cessation on airway inflammation in COPD and
525 asymptomatic smokers. *Eur Respir J* 2005; **26**(5): 835-45.

- 526 27. Hashemi-Bajgani SM, Abbasi F, Shafahi A, Yazdani R, Samareh Fekri M.
527 Association of Bile Acid and Pepsin Micro-aspiration with Chronic Obstructive
528 Pulmonary Disease Exacerbation. *Tanaffos* 2019; **18**(1): 52-7.
- 529 28. Pragman AA, Lyu T, Baller JA, et al. The lung tissue microbiota of mild and
530 moderate chronic obstructive pulmonary disease. *Microbiome* 2018; **6**(1): 7.
- 531 29. Bouquet J, Tabor DE, Silver JS, et al. Microbial burden and viral exacerbations
532 in a longitudinal multicenter COPD cohort. *Respir Res* 2020; **21**(1): 77.
- 533 30. Saladie M, Caparros-Martin JA, Agudelo-Romero P, Wark PAB, Stick SM,
534 O'Gara F. Microbiomic Analysis on Low Abundant Respiratory Biomass Samples;
535 Improved Recovery of Microbial DNA From Bronchoalveolar Lavage Fluid. *Front*
536 *Microbiol* 2020; **11**: 572504.
- 537 31. Klindworth A, Pruesse E, Schweer T, et al. Evaluation of general 16S ribosomal
538 RNA gene PCR primers for classical and next-generation sequencing-based diversity
539 studies. *Nucleic Acids Res* 2013; **41**(1): e1.
- 540 32. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database
541 project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;
542 **41**(Database issue): D590-6.
- 543 33. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple
544 statistical identification and removal of contaminant sequences in marker-gene and
545 metagenomics data. *Microbiome* 2018; **6**(1): 226.
- 546 34. Caparros-Martin JA, Lareu RR, Ramsay JP, et al. Statin therapy causes gut
547 dysbiosis in mice through a PXR-dependent mechanism. *Microbiome* 2017; **5**(1): 95.
- 548 35. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative
549 models for microbial metagenomics. *PLoS One* 2012; **7**(2): e30126.
- 550

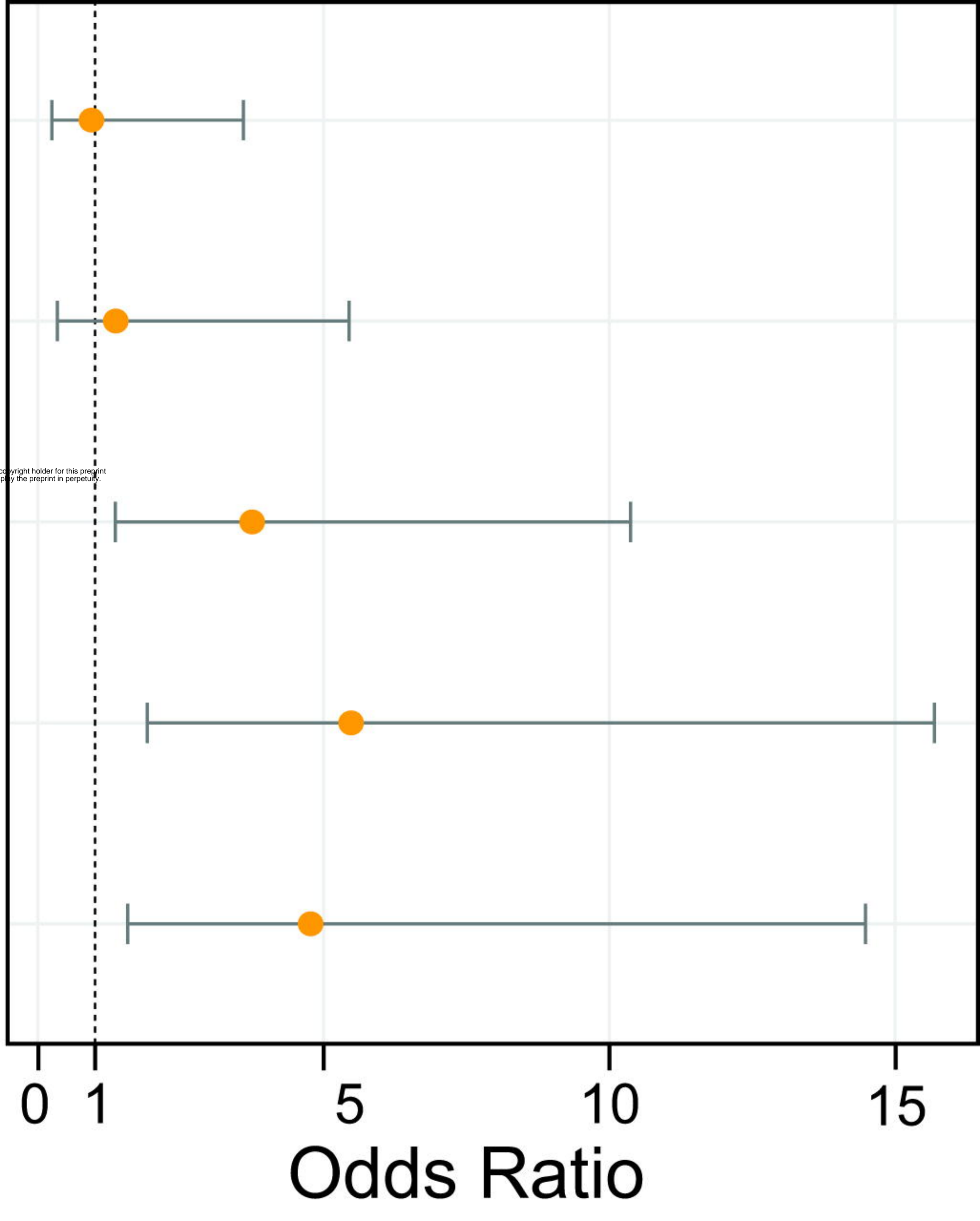
Healthy smokers

GOLD 4

GOLD 3

GOLD 2

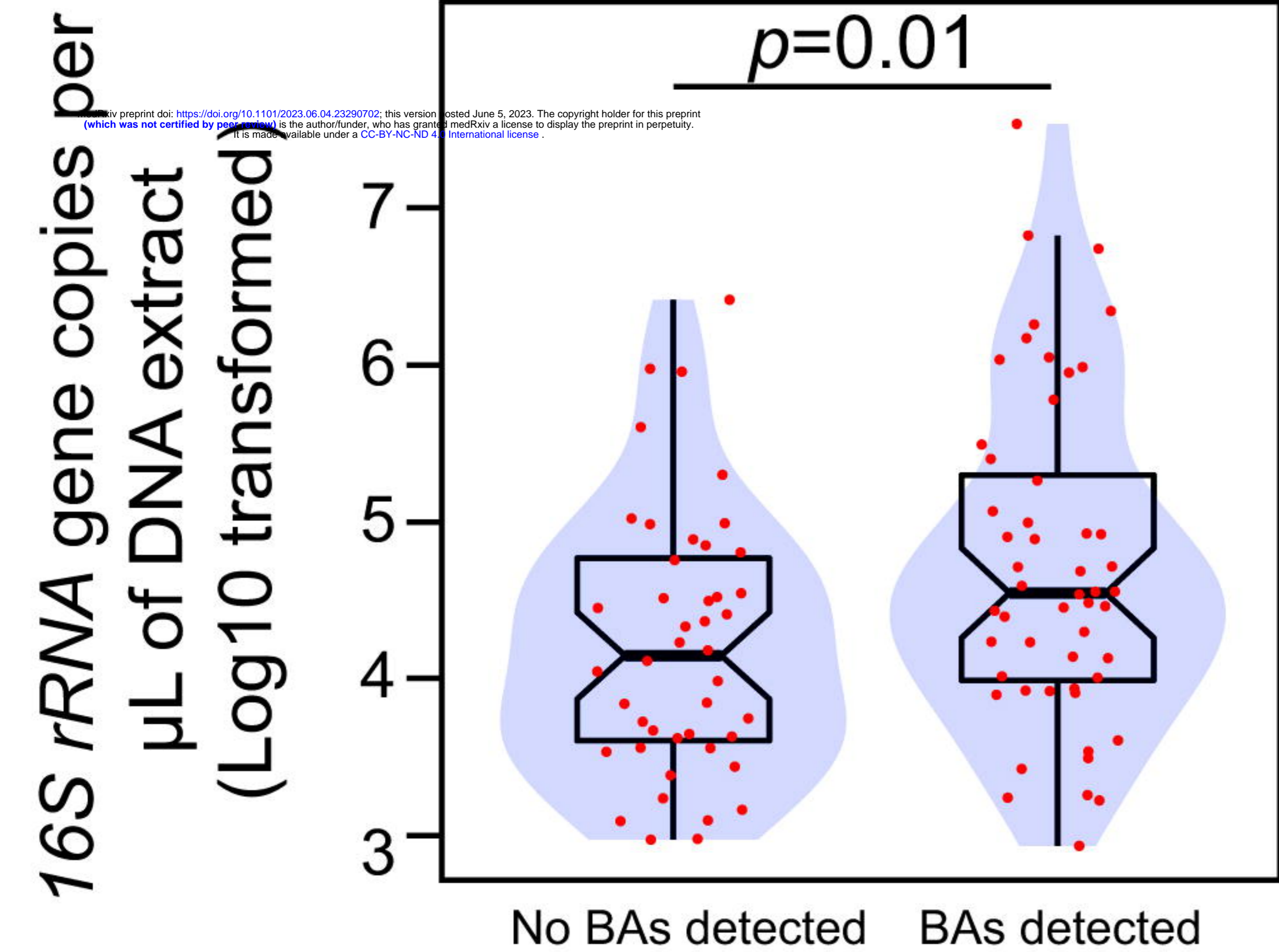
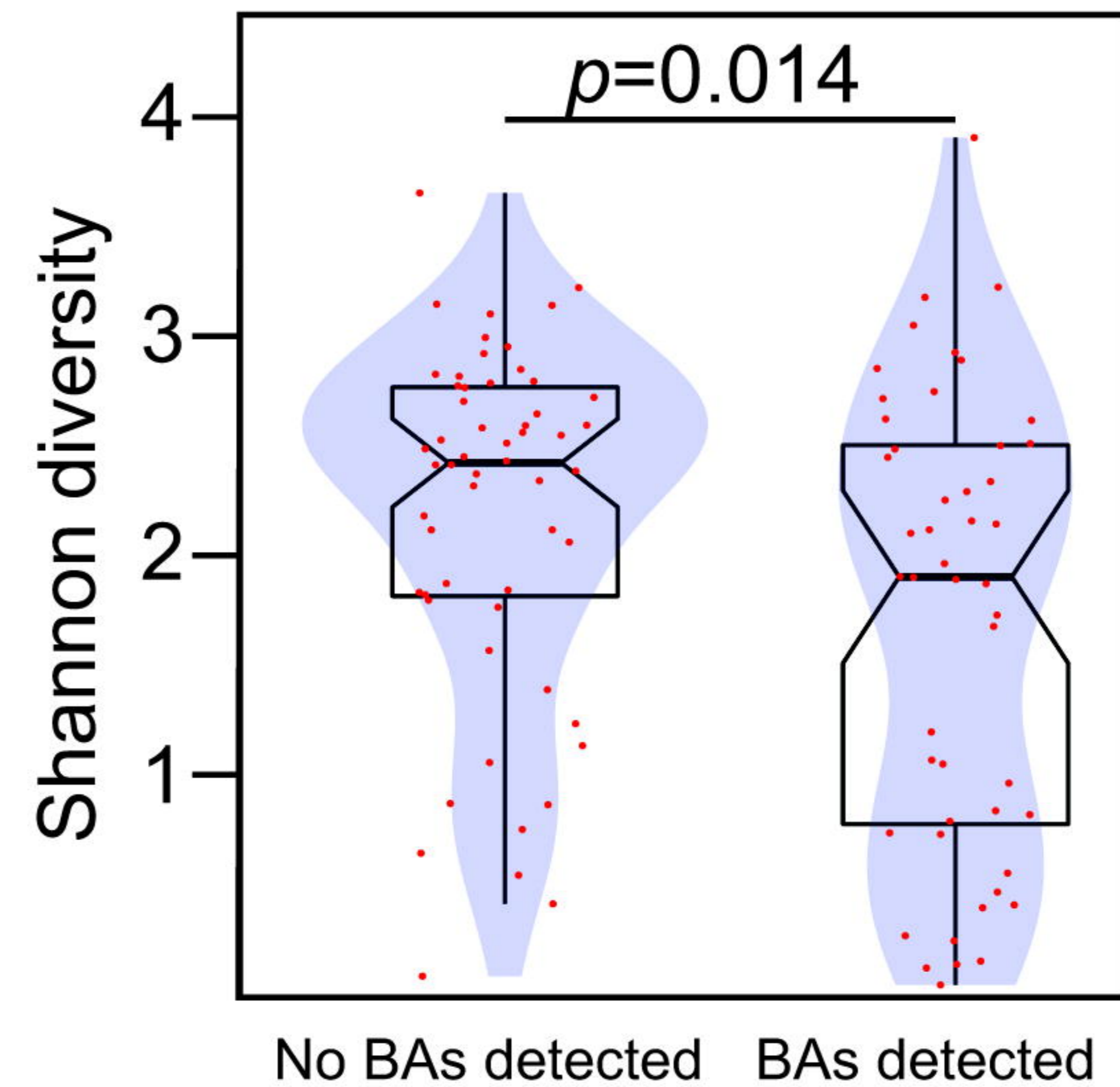
GOLD 1



0 1 5 10 15

Odds Ratio

medRxiv preprint doi: <https://doi.org/10.1101/2023.06.04.23290702>; this version posted June 5, 2023. 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. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

A**B****C**