Endotyping Chronic Obstructive Pulmonary Disease, Bronchiectasis, and the "Chronic Obstructive Pulmonary Disease–Bronchiectasis Association"

Huang, Jeffrey T.-J.; Cant, Erin; Keir, Holly R.; Barton, Alun K.; Kuzmanova, Elena; Shuttleworth, Morven

Published in:
American Journal of Respiratory and Critical Care Medicine

DOI:
10.1164/rccm.202108-1943OC

Publication date:
2022

Document Version
Peer reviewed version

Originally Published in: Huang, J. T-J., Cant, E., Keir, H. R., Barton, A. K., Kuzmanova, E., Shuttleworth, M., Pollock, J., Finch, S., Polverino, E., Bottier, M., Dicker, A. J., Shoemark, A., & Chalmers, J. D. (2022). Endotyping Chronic Obstructive Pulmonary Disease, Bronchiectasis, and the "Chronic Obstructive Pulmonary Disease–Bronchiectasis Association". American Journal of Respiratory and Critical Care Medicine, 206(4), 417-426. https://doi.org/10.1164/rccm.202108-1943OC

The final publication is available at https://doi.org/10.1164/rccm.202108-1943OC

Endotyping COPD, bronchiectasis and the 'COPD-bronchiectasis association'

Jeffrey T-J Huang[1], Erin Cant[2], Holly R Keir[2], Alun K Barton[1], Elena Kuzmanova[1], Morven Shuttleworth[2], Jennifer Pollock[2], Simon Finch[2], Eva Polverino[3], Mathieu Bottier[2], Alison J Dicker[1], Amelia Shoemark[2], James D Chalmers[2]

1. Division of Systems Medicine, School of Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, DD1 9SY
2. Division of Molecular and Clinical Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, DD1 9SY
3. Pneumology Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain.

Corresponding Authors: James D Chalmers, Division of Molecular and Clinical Medicine, University of Dundee, Dundee, DD1 9SY, UK, jchalmers@dundee.ac.uk, phone 01382383642;

Word count: 3724 (Max = 3500 words)

Keywords: COPD, bronchiectasis, the 'COPD-bronchiectasis association', endotype, proteome, microbiome, sputum

Running title: Endotyping COPD, bronchiectasis and the 'COPD-bronchiectasis association'

Acknowledgements: Funded by the European Respiratory Society through the EMBARC2 consortium. EMBARC2 is supported by project partners AstraZeneca, Chiesi, Grifols, Insmed, Janssen, Novartis and Zambon. JDC is supported by the British Lung Foundation Chair of Respiratory Research and a Scottish Senior Fellowship from the Chief Scientist Office.

Conception and design: JTJH, HRK, EP, AD, AS, JDC
Data collection and laboratory work: EC, HRK, AKB, EK, MS, JP, SF, MB, AJD, AS, JDC

Analysis and interpretation: JTJH, EC, HRK, AJD, AS, JDC
Drafting the manuscript: JTJH, JDC
Reviewing the manuscript for important intellectual content and approving the final version: all authors

At a Glance Commentary

Scientific knowledge on the subject: Chronic obstructive pulmonary disease (COPD) and bronchiectasis are two diseases with overlapping clinical presentation. Co-diagnosis of both diseases commonly occurs in some patients (termed as the 'COPD-bronchiectasis association') but the mechanisms, risk factors and potential management options for patients with the 'COPD-bronchiectasis association' are largely unknown. We hypothesized that the 'COPD-bronchiectasis association' syndrome would have an underlying pathophysiology different from COPD but similar to bronchiectasis, and the resulting molecular and microbial features would produce biologically informed patient classification.

What this study adds to the field: We demonstrate for the first time that patients with COPD and the 'COPD-bronchiectasis association' presented different profiles in their lung microbiota and host responses, and that the underlying pathophysiology of the 'COPD-bronchiectasis association' is closer to that of bronchiectasis. These results were validated in an independent cohort showing that neutrophilic inflammation, increased abundance of pathogenic proteobacteria and dysregulation of mucins are key processes associated with bronchiectasis and the COPD-bronchiectasis association. We propose here a biologically informed patient classification for airways disease patients according to their clinical, sputum microbiome and protein profiles.

Abstract

Rationale

Bronchiectasis and chronic obstructive pulmonary disease (COPD) are two disease entities with overlapped clinical features and co-diagnosis frequently occurs (termed as the 'COPD-bronchiectasis association').

Objectives

To investigate the sputum microbiome and proteome in patients with bronchiectasis, COPD, and the 'COPD-bronchiectasis association' with the aim of identifying endotypes that may inform treatment.

Methods

Sputum microbiome and protein profiling were carried out using 16S rRNA amplicon sequencing and a label-free proteomics workflow, respectively, in a cohort comprising patients with COPD (n=43), bronchiectasis (n=30) and the 'COPD-bronchiectasis association' (n=48). Results were validated in an independent cohort of 91 patients (n=28-31 each group) using targeted measurements of inflammatory markers, mucins and bacterial culture.

Measurements and main results

Principal component analysis of sputum microbiome and protein profiles showed a partial separation between the COPD and the 'COPD-bronchiectasis association' group. Further analyses revealed that patients with the 'COPD-bronchiectasis association' had a higher abundance of proteobacteria, higher expression of mucin-5AC and proteins from the "neutrophil degranulation" pathway compared to those with COPD. In contrast, COPD patients had an elevated expression of mucin-5B and several peptidase inhibitors, higher abundance of common commensal taxa, and a greater microbiome diversity. The profiles of 'COPD-bronchiectasis association' and bronchiectasis groups were largely overlapping. Five endotypes were proposed with differential inflammatory, mucin and microbiological features. The key features related to the 'COPD-bronchiectasis association' were validated in an independent cohort.

Conclusion

Neutrophilic inflammation, differential mucin expression and Gram-negative infection are dominant traits in patients with the 'COPD-bronchiectasis association'.

Word count: 244 (limit =250)

Introduction

Chronic obstructive pulmonary disease (COPD) and bronchiectasis are two diseases with overlapping clinical presentation including cough, sputum production, breathlessness and increased susceptibility to exacerbations. Since bronchiectasis and COPD are defined by different criteria - clinical and radiological features for bronchiectasis and lung function and a history of relevant exposures for COPD, co-diagnosis of both diseases occurs in some patients (termed as the 'COPD-bronchiectasis association' or "overlap" (1)). Studies have shown that the prevalence of bronchiectasis defined by radiological evidence on computed tomography in patients with COPD is highly variable ranging from 4% to 72% depending on the criteria of bronchiectasis and study populations.(2) Conversely, around 20% of COPD were reported as an underlying diagnosis in both the US and European bronchiectasis registries.(3, 4)

The COPD-bronchiectasis overlap syndrome has recently been recognised as a neglected area of research due to the lack of understanding of pathophysiology of this syndrome and the absence of guidelines for clinical practice.(5) Current evidence indicates that the presence of bronchiectasis in COPD is associated with greater disease burden and worse outcomes including a higher frequency of *Pseudomonas aeruginosa* infection (6), more frequent exacerbations (7), longer duration of hospitalisation (8), higher disease severity (7) and all-cause mortality (9-11). However, the mechanisms, risk factors and potential management options for patients with the 'COPD-bronchiectasis association' are mostly unknown.

Understanding underlying molecular and microbial endotypes and clinical phenotypes may help to guide treatment approaches for heterogenous conditions such as the 'COPD-bronchiectasis association' (12). "Treatable traits" have been proposed as a new concept for the management of complex airway diseases by attempting to move beyond disease labels.(13, 14) The essence of this treatable traits approach is to identify biomarkers to allow targeted treatment. Currently the underlying biology of the 'COPD-bronchiectasis association' is poorly understood meaning that it is difficult to apply a treatable traits approached to find targeted therapies.

A systems biology approach using multi-omics techniques may provide a holistic view to disease pathophysiology (15) and allow the search for biomarkers that are associated with certain phenotypes or endotypes. Biomarkers to guide the treatable traits approach can be

derived from either the host, through approaches such as proteomics, or from airway microbial communities through microbiome sequencing. By combining sputum microbiome and proteome information with clinical phenotypic data, we have recently found that a reduction in microbiome diversity and an increase in neutrophil extracellular traps proteins are associated with greater disease severity, exacerbation frequency and severe exacerbations, as well as a higher risk of mortality in patients with bronchiectasis (16) and COPD (16, 17). In this study, we took a similar holistic approach to investigate the endotypic differences between patients with COPD, the 'COPD-bronchiectasis association' and bronchiectasis, followed by a validation of key features in an independent cohort. We hypothesized that the 'COPD-bronchiectasis association' syndrome would have an underlying pathophysiology different from COPD but similar to bronchiectasis, and the resulting molecular and microbial features would produce biologically informed patient classification.

Methods

Study design

Patients were enrolled from two separate cohorts which were combined for analysis. Bronchiectasis subjects were enrolled from the Tayside Bronchiectasis Registry Integrating Datasets, Genomics and Enrolment into Clinical Trials (TAYBRIDGE) observational study (18) and patients with COPD enrolled from the Tayside Allergy and Respiratory Disease Information System (TARDIS) registry study (19). Both of these studies allowed patients with overlapping COPD and bronchiectasis to be enrolled.

Bronchiectasis patients had to be ≥18 years, have a high definition CT confirmed diagnosis of bronchiectasis and have clinical symptoms (cough, sputum production, dyspnoea or respiratory infections) consistent with bronchiectasis to be included. Patients were excluded if there was an inability to give informed consent or had active tuberculosis or lung cancer. Patients with cystic fibrosis or pulmonary fibrosis with secondary bronchiectasis were also excluded.

Diagnosis of radiological bronchiectasis in the bronchiectasis and overlap groups was made using the Fleischner society criteria (20) requiring a bronchial:arterial ratio >1.0, lack of tapering or visible airways within 1cm of the pleural surface.

COPD patients were included if >40 years; if they had a FEV_1/FVC ratio <70% at a screening visit, at least a 10-pack year smoking history and had a clinical diagnosis of COPD. Exclusion criteria included the inability to give informed consent; primary diagnosis of asthma; and systematic immunosuppression (excluding prednisolone at 5mg or less daily).
Patients in the bronchiectasis, COPD and bronchiectasis-COPD association groups all needed to be clinically stable and free of antibiotic or oral corticosteroid therapy for 4 weeks prior to enrolment. All relevant medical history (comorbidities, current medications, significant past conditions, operations and diagnostic procedures) were recorded at screening.

Patients meeting the inclusion criteria for both bronchiectasis and COPD groups indicating the presence of bronchial dilatation on HRCT, clinical symptoms, airway flow obstruction and a smoking history were classified as "the 'COPD-bronchiectasis association'". To be included in the "COPD only" cohort, participants therefore required a CT scan in the previous 5 years demonstrating no evidence of bronchiectasis. To be included in the bronchiectasis cohort, participants had an absence of a clinical COPD diagnosis and an absence of a relevant smoking history (<10 packs/year) and/or the absence of airflow obstruction on spirometry performed as screening.

To validate findings in the initial patient cohort, a further 91 patients were enrolled from the same cohorts meeting the same inclusion and exclusion criteria with targeted measurement of inflammatory markers, mucin measurement and bacterial culture as described below.

Patients gave written informed consent and the studies were approved by the East of Scotland Research Ethics committee 12/ES/0059 and 13/ES/0030.

Clinical Assessment and Sampling

Patients provided spontaneous sputum samples at enrolment, when clinically stable (defined as no antibiotics, apart from normal prophylactic antibiotics, in the previous 4 weeks). Clinical

and quality of life assessments were carried out as described previously (18). Severity of disease was evaluated using the multidimensional bronchiectasis severity index (BSI)(21) for bronchiectasis and the GOLD criteria (22) for COPD. Symptoms were evaluated using the quality of life bronchiectasis respiratory symptom scale (QOLBrss)(23), a validated instrument in bronchiectasis and the St Georges Respiratory Questionnaire (SGRQ) in COPD. The radiological severity of bronchiectasis was scored using the modified Reiff scale, which provided 1-3 points per lobe affected by bronchiectasis according to whether bronchiectasis is cylindrical (1), varicose (2) or cystic (3) as previously described (24).

Cohort matching

Bronchiectasis is more common in females and COPD is more common in males in many cohorts. Lung function and radiological parameters are different between the groups and the bronchiectasis/COPD overlap syndrome is associated with greater disease severity than COPD or bronchiectasis alone. It was therefore determined a-priori not to attempt to match the cohorts, which would select a non-representative sample of these diseases. A flow chart of patients included in the current study is shown in Suppl. Figure 1.

Sputum proteomics

Protein profiling of sputum supernatant was performed using a label-free proteomics workflow nano-flow (see Supplementary File).

Sputum Microbiome

DNA was extracted from sputum and the V3 and V4 region of the bacterial 16S rRNA gene sequenced as previously described (16). Alpha diversity was measured by determining the Shannon-Wiener Diversity (SWDI). Further details are provided in the online supplement. In the validation cohort, sputum was sent for standard bacterial culture.

Targeted analysis of sputum proteins

Neutrophil elastase activity was measured using an activity-based immunoassay (ProteaseTag, ProAxsis Ltd, Belfast, UK). MUC5B and MUC5AC were measured by stable isotope dilution LC-MS/MS method (see Supplementary file). A panel of inflammatory

cytokines were measured by a multiplex assay (MSD, Maryland, US), Interleukin-8 was measured by ELISA (R+D systems) and sputum DNA was measured by Picogreen assay.

Statistical Analysis

Chi-square analysis with Yates correction was performed using SPSS (version 22.0, IBM). Protein identification and label-free quantification were carried out using Maxquant (version 1.4.1.2) against Uniprot-human database (version 2014-07-09). False discovery rate (FDR) for protein identification was set to 1% at protein level. Principal component analysis and partial least square analysis (PLS-DA) were carried out using SIMCA-P (version 13.0.3, UMETRICS). To avoid overfitting in PLS-DA models, procedures including PCA (25), random permutation of dataset rows (26) and CV-ANOVA(27) were used. For statistical analysis, the dataset was log2 transformed before subjecting to t test using Perseus (version 1.5.4.1). A permutation-based FDR method (9) was used and corrected p values of $p<0.05$ is considered significant.

Results

Differences in sputum microbiome and proteome among COPD, bronchiectasis and the association

We carried out protein profiling and microbiome analysis in 43 patients with COPD, 30 patients with bronchiectasis, and 48 patients with the 'COPD-bronchiectasis overlap syndrome'. Demographic characteristics of the patients are shown in **Table 1**. Among the bronchiectasis participants 18 were classified as idiopathic, 5 post-infective, 5 had connective tissue diseases and 2 immunodeficiency. The participants in the COPD and the 'COPD-bronchiectasis association' groups had similar age, smoking status, FEV_1, blood eosinophil counts, MRC dyspnea score and quality of life scores. The COPD group had a higher proportion of male participants and lower FEV_1/FVC ratio as expected in the disease population. The use of inhaled corticosteroid and oral antibiotics in the previous year were higher in the 'COPD-bronchiectasis association' group. Between the bronchiectasis and the 'COPD-bronchiectasis association' groups, the participants in the bronchiectasis had a lower bronchiectasis severity and higher lung function indices, whereas the age and gender distribution were similar.

Sputum microbiome and protein profiling were carried out using 16S rRNA gene amplicon sequencing and a label-free proteomics workflow, respectively. After quality control, a total of 741 variables comprising 127 variables from microbiome (120 at the genus level, 6 at the phylum level and SWDI) and 614 variables from proteome were included in the analysis. The abundance of microbiome at the phylum level and microbiome diversity were compared between the groups. Patients with the COPD-bronchiectasis association had a higher relative abundance of proteobacteria (p=0.001) and lower abundance of Bacteroidetes (p=0.002), Firmicutes (p<0.05) and other phyla (p<0.001) (**figure 1A**), as well as a lower SWDI (p<0.001)(**figure 1B**) compared to those with COPD alone. The significance for *Proteobacteria*, *Bacteroidetes* and other phyla remained after adjusted for previous use of oral antibiotic, inhaled antibiotics, ICS and gender. No clear differences were observed between the bronchiectasis and the 'COPD-bronchiectasis association' groups.

Principal component analysis (PCA) using all 741 variables showed a partial separation between the COPD and the 'COPD-bronchiectasis association' groups (**figure 2A**), with the bronchiectasis and the 'COPD-bronchiectasis association' groups largely overlapped (**figure 2A**). The partial separation was driven by higher abundance of *Proteobacteria* and lower abundance of *Firmicutes* and *Bacteroidetes* and a higher SWDI, as well as increased expression of lymphocyte cytosolic protein 1 (LCP1), catalase (CAT), metalloproteinase 9 (MMP-9), and cathelicidin antimicrobial peptide (CAMP), and lower expression of tissue inhibitor of metalloproteinase 1 (TIMP1), polymeric immunoglobulin receptor (PIGR), and immunoglobulin kappa chain (IgK@) (**figure 2B**). Excluding the phylum variables and SWDI did not affect the model (**Suppl figure S-2**). These data implied that the underlying pathophysiology of COPD and the 'COPD-bronchiectasis association' may be different.

To further examine this possibility, we carried out partial least square discriminant analysis (PLS-DA) which displays the largest difference between the two groups with all variables included. The scores plots showed an almost complete separation of the COPD patients to those with the 'COPD-bronchiectasis association' (**figure 3A**). The loadings plot indicated that the participants with the 'COPD-bronchiectasis association' had higher protein expression of Mucin-5AC (MUC5AC), orosomucoid 1 (ORM1), bactericidal/permeability-increasing protein-like 1 (BPIFB1), myeloperoxidases (MPO), and neutrophil elastase (ELANE) (**figure 3B**). In

contrast, COPD patients without bronchiectasis had an elevated expression of MUC5B, IgK@ and higher abundance of *Fusobacterium and Veillonella,* which are common commensal taxa (**figure 3B**). As expected, higher abundance of Proteobacteria and lower other phyla and SWDI were associated with the COPD-bronchiectasis association (**figure 3B**). These observations were confirmed by t test with FDR correction (**Suppl. table S-1**). Similar results were found when excluding those who had long-term oral antibiotics use in the previous year (**Suppl. figure S-4**) and similarly, gender did not appear to have a significant effect (**Suppl. figure S-5)**. Pathway analysis of differentially up-regulated proteins in the 'COPD-bronchiectasis association' group revealed an over-representation of the "neutrophil degranulation" pathway (p=2.8E-4, corrected by the Bonferroni method) whereas the down-regulated proteins were over-represented with "peptidase inhibitors". Taken together with the data from microbiome diversity and distribution at the phylum level, these results indicate that the presence of bronchiectasis in COPD is associated with proteobacteria dysbiosis, neutrophilic inflammation, lower expression of peptidase inhibitors and differential expression of MUC5B/MUC5AC.

Between bronchiectasis and the COPD-bronchiectasis association' groups, a lower expression of GSTP1 in the bronchiectasis group was the only noticeable difference (corrected $p<0.01$, t test). No clear differences were observed among the first seven components of PCA (data not shown). This is further confirmed by projecting data from patients with bronchiectasis without COPD to the original PLS model shown in figure 2B. The data showed that bronchiectasis and the 'COPD-bronchiectasis association' shared similar sputum proteome and microbiome profiles (**Suppl. figure S-6**).

Validation

The main findings were validated in an independent cohort of 32 patients with COPD, 31 patients with bronchiectasis and 28 patients with the 'bronchiectasis-COPD association'. The characteristics of the patients are shown in **table 2**. Consistent with the original study, we found dysregulation of mucins with an increase of MUC5AC/MUC5B ratio in bronchiectasis (p=0.02) and the bronchiectasis-COPD association (p=0.02) compared to COPD alone (**supplementary figure S-7A**). Neutrophil elastase activity was also significantly higher in

bronchiectasis and the association compared to COPD alone (p≤0.001, **supplementary figure S-7B**). Gram-negative bacterial infection, predominantly *P. aeruginosa* and *H. influenzae* infection were more frequent in bronchiectasis and the COPD-bronchiectasis association' compared to COPD (**supplementary figure S-7C**).

PCA was repeated integrating the targeted and inflammatory and microbial data in the validation cohort. The PCA plot in **figure 4A** shows a clear separation between COPD and bronchiectasis or the COPD-bronchiectasis association' group. No obvious effect from the previous use of oral antibiotics was observed (**Suppl figure S-8**). The loadings plot in **figure 4B** confirms the key features of the endotypes with bronchiectasis being most strongly associated with MUC5AC, DNA, neutrophil elastase and IL-8. Predominant COPD was associated with increased MUC5B.

Endotypes based on sputum microbiome and proteome

Several key features revealed in the PCA and PLS-DA may be useful for identifying potential endotypes. The most dominant features observed were neutrophilic inflammation, mucin secretion, microbiome diversity and proteobacteria described above. Another feature is shown in the second component of the PCA (**figure 2**), which identified a small cluster of patients with high levels of actin and actin interacting proteins (ACTG1, EZR, AHSG) including gelsolin, the release of which is regulated by interleukin-4(28). Hypothesising that Th2 inflammation may be related to this cluster we observed clustering of eosinophil peroxidase (EPX) and galectin-10 (CLC) in these patients (**figure 2B**). This Th2 endotype was further confirmed in the validation cohort where a cluster of patients with predominant Th2 inflammation reflected by increased interleukin-4, interleukin-5, thymic stromal lymphopoietin (TSLP), thymus and activation-regulated chemokine (TARC) and eotaxins (**figure 4**). This endotype was observed in a subset of COPD, bronchiectasis and overlap patients. Additionally, features that drive inter-patient variability as shown in **figure 3** would provide further subtyping information. These include two groups of variables with opposing influences: one with a cluster of proteins including metalloproteinase-8, MMP-9, LCP1 and *Haemophilus* relative abundance, and the other group with SLPI, PIGR and IgJ (**figure 3B**). Overall, these key features can be plotted in 3-dimensional space with x axis driving predominantly by neutrophilic inflammation and microbiome diversity, y axis by innate

immune activation, proteases and *Haemophilus*, and z axis by eosinophil inflammation (**figure 5**).

As many of these features are known to be of prognostic value in COPD or bronchiectasis (see Discussion), we proposed five endotypes based on these key features (**figure 5**). Firstly, an endotype we refer to as diverse and protective, associated with high microbial diversity and commensal taxa, and increased MUC5B, protease inhibitors and immunoglobulins. A second endotype was termed "*Haemophilus* proteolytic" with *Haemophilus* in the microbiome and high levels of MMP8/MMP9 and low MUC5AC/MUC5B. Endotype III is associated with high levels of IgJ, PIGR, SLPI, and MUC5AC/MUC5B and shows low microbial diversity with organisms such as Stenotrophomonas and other proteobacteria and is referred to as the "infected, epithelial response" endotype. Endotype IV is dominated by neutrophil extracellular trap proteins and Proteobacteria abundance. Endotype V is driven by Th2 inflammation.

Discussion

The results from this study demonstrated for the first time that patients with COPD and the 'COPD-bronchiectasis association' presented different profiles in their lung microbiota and host responses. Importantly, our data revealed that the underlying pathophysiology of the 'COPD-bronchiectasis association' is closer to that of bronchiectasis. Based on these results, we propose to classify patients with COPD, bronchiectasis, or the 'COPD-bronchiectasis association' into four endotypes according to their sputum microbiome and protein profiles which show evidence of different underlying host response and microbial profiles. Importantly, the distinction between these groups was not evident from clinical features or disease labels alone indicating that disease labels alone do not accurately identify potentially treatable mechanisms of disease.

These endotypes detected in a clinical stable state may represent potential "treatable traits" that could align with current treatment options in COPD. The diverse-protective endotype are likely to have the best prognosis (29) and may not require anti-inflammatory or antimicrobial treatment. The Haemophilus-proteolytic endotype is associated with *Haemophilus* infection

with a higher MMP-8/MMP-9 driven inflammatory state so antibiotic therapy targeting *Haemophilus* and anti-inflammatory treatment (e.g. PDE4 inhibitors) could be considered accordingly. The identification of this endotype is intriguing as tetracyclines are effective against *Haemophilus* and have activity that blocks matrix metalloproteinase activity, which may therefore be a theoretical option in this group. Tetracyclines are currently being tested in COPD trials (NCT02305940). The infected-epithelial response endotype has features of bronchiectasis, Gram-negative infection and MUC5AC overexpression. We speculate that such patients may benefit from macrolide treatment which has been shown to be effective against *P. aeruginosa* infection (30), and macrolides have also been shown to inhibit MUC5AC release from epithelial cells independently of their antibiotic effect(31). The proteobacteria-neutrophilic endotype also has a feature of bronchiectasis with increased MUC5AC and features of excessive neutrophil activation, which we recently linked to the formation of neutrophil extracellular traps (32). We have recently shown that macrolides may also be beneficial in this group (30). In addition, recent evidence shows that this mechanism may be targeted in future with dipeptidyl peptidase-1 inhibition which was found to prolong the time to first exacerbation in bronchiectasis patients (33). The Th2 endotype may benefit may benefit from inhaled corticosteroids or other treatments targeting Th2 inflammation if they experience frequent exacerbations. These proposals are hypothetical and need to be tested in appropriately controlled clinical studies. Nevertheless, they demonstrate the potential value of identifying linked mechanisms of disease.

Several endotypic biomarkers proposed in this study have been shown to provide prognostic information in the context of COPD and bronchiectasis although they have not been reported in the COPD-bronchiectasis overlap syndrome. For example, *Pseudomonas* and *Haemophilus* infections have been associated with increased risk of exacerbation and death in COPD. (34, 35) In bronchiectasis, *Pseudomonas* infections have been associated with more hospital admissions (36, 37), worse quality of life (36, 37), and a rapid decline in lung function (38) compared to those with infections with other bacteria. Recently, we have also reported that reduced sputum microbiome diversity, associated with Proteobacteria (predominantly *Haemophilus*) dysbiosis is associated neutrophilic inflammation and a significantly increased risk of mortality in COPD (16). In bronchiectasis, *Pseudomonas*- and *Haemophilus*-dominated microbiomes have been shown to be linked to severe disease and frequent exacerbations

(39). Likewise, the sputum protein biomarkers including MPO (40), ELANE (18), MUC5AC(41), MUC5B(41), CAMP (or LL-37) (42, 43), and SLPI (44) have been shown to be pathologically relevant and prognostically important in COPD and/or bronchiectasis. For example, sputum MPO levels were higher in COPD than controls and further pronounced during exacerbations (40). COPD patients with more frequent reported exacerbations had lower sputum concentrations of SLPI (44). Circulating antimicrobial peptide LL-37 levels was shown to associated with high risk of frequent exacerbations in COPD (43). These previous reports strongly support the relevance of using these biomarkers for endotyping patients with COPD, bronchiectasis and the 'COPD-bronchiectasis association'.

One of the striking observations in this study is the contrasting expression patterns of mucins, where MUC5AC was over-expressed in patients with the 'COPD-bronchiectasis association' whereas MUC5B was up-regulated in COPD patients without bronchiectasis. MUC5AC is found predominantly in the goblet cells in the surface epithelia throughout the central conducting airways, whereas MUC5B is found mainly in submucosal glands of central airways (45). Evidence from animal studies indicated that MUC5B but not MUC5AC is a dominant regulator of homeostatic microbial elimination (i.e. mucocilary clearance) in the airway (46). Loss of MUC5B inhibits innate inflammatory responses with suppressed interleukin-23 production resulting in an accumulation of alveolar macrophages with impaired ability to phagocytose and clear *Staphylococcus*.(47) Therefore, a lower MUC5B expression in the 'COPD-bronchiectasis association' group implies that these patients may have a compromised mucocilary clearance. In contrast, MUC5AC is known to confer resistance to viral infection (48) but is detrimental for allergic airway hyperreactivity (49) and acute lung injury (50) with enhanced neutrophil trafficking and inflammation. Similar to the latter, the over-expressed MUC5AC in the 'COPD-bronchiectasis association' observed in this study was associated with neutrophilic inflammation, which is likely to have a worse outcome as seen in patients with bronchiectasis in our previous studies.(18, 44) A recent study from the SPIROMICS consortium also showed a striking increase in MUC5AC concentrations with COPD disease severity.(41) The increase has been shown to associated with disease progression. (51)

The study has several limitations. First, this is a cross-sectional study and the study was not designed or powered to demonstrate the prognostic implications of these endotypes. However, some of biomarkers (e.g. ELANE, CAMP, SLPI, MMP-9) described have been shown

to provide prognostic values in bronchiectasis and COPD. Second, the demographics of patients with bronchiectasis were not well matched to COPD patients, as expected. However, the PLS-DA analysis employed has taken into account the contributions of those demographic variables and various sensitivity analyses have shown that patient characteristics such as sex and age are not associated with sputum microbiome or protein profiles in this population. Last, the current study focused on data collected during a clinically stable state so the results may not be extrapolated to patients during exacerbations.

In summary, our study indicates that the presence of bronchiectasis, lung microbiota and differential host responses are important determinants for endotyping patients with COPD, and the 'COPD-bronchiectasis association'. This information may be used for biologically informed patient classification of COPD and the 'COPD-bronchiectasis association'.

Table 1. Demographic details of study participants in the first cohort.

	COPD	The 'COPD-bronchiectasis association'	Bronchiectasis	p value (COPD vs the 'COPD-bronchiectasis association')	p value (the 'COPD-bronchiectasis association' vs bronchiectasis)
N	43	48	30		
Age	73 (70-76)	72 (63-81)	72 (69-74)	n.s.	n.s.
Gender (m/f)	31/12	24/24	15/15	0.03	n.s.
Smoking status (never-/ex-/current)	0/7/36	0/6/42	19/10/1	n.s.	<0.0001
Smoking (pack year)	44 (34-54)	36(32-39)	17 (7-27)	n.s.	<0.0001
Body mass index	29 (28-31)	27 (26-29)	28 (26-30)	n.s.	n.s.
FEV_1 (litres)	1.6 (1.4-1.8)	1.4(1.2-1.6)	2.0 (1.8-2.2)	n.s.	<0.0001
FEV_1 (% pred)	69 (62-75)	64 (57-71)	85 (76-93)	n.s.	<0.0001
FVC (Litres)	3.1 (2.8-3.4)	2.6 (2.3-2.9)	3.0 (2.6-3.3)	0.03	n.s
FEV_1/FVC ratio	51 (48-54)	54 (51-57)	67 (61-72)	0.01	<0.0001
Blood eosinophil (%)	0.2 (0.16-0.24)	0.17 (0.13-0.20)	0.17 (0.13-0.22)	n.s.	n.s.
MRC dyspnoea score	2.2 (1.8-2.6)	2.2 (1.9-2.6)	1.4 (0.9-1.9)	n.s.	0.02
Exacerbation frequency*	2.7 (1.9-3.4)	2.8 (2.2-3.4)	2.0 (1.3-2.7)	n.s.	n.s.
SGRQ	51 (44-58)	53 (46-60)	n.a.	n.s.	-
QoL-B RSS	NA	56 (51-62)	63 (56-71)	-	n.s.
CAT	20 (18-23)	22 (20-25)	n.a.	n.s.	-
Bronchiectasis severity index	n/a	8.2 (7.1-9.3)	6.9 (5.4-8.5)	-	n.s.
ICS use*	67.4%	77.1%	50.0%	n.s.	0.01
Long term Oral antibiotics use*	4.7%	50.0%	34.2%	<0.0001	n.s.
Inhaled antibiotics use*	0.0%	8.3%	5.3%	n.s	n.s.

Data: Mean (95% confidence interval); n.s: not significant; n.a.: not available

Table 2. Demographic details of study participants in the validation cohort.

	COPD	The 'COPD-bronchiectasis association'	Bronchiectasis	p value (COPD vs the 'COPD-bronchiectasis association')	p value (the 'COPD-bronchiectasis association' vs bronchiectasis)
N	32	28	31		
Age	62 (59-64)	72 (69-75)	70 (66-74)	<0.001	n.s.
Gender (m/f)	23/9	18/10	18/13	n.s.	n.s.
Smoking status (never-/ex-/current)	(0/23/9)	(0/24/4)	(20/10/1)	n.s.	<0.0001
Smoking (pack year)	39 (31-47)	45 (38-51)	5 (0-9)	n.s.	<0.0001
Body mass index	25 (23-27)	27 (25-39)	29 (26-32)	n.s.	n.s.
FEV_1 (litres)	1.2 (1.0-1.4)	1.4 (1.2-1.6)	1.9 (1.3-2.2)	n.s.	0.006
FEV_1 (% pred)	44 (38-50)	62 (54-70)	76 (67-85)	0.005	0.02
FVC (Litres)				n.s.	0.05
FEV_1/FVC %	38 (34-42)	49 (44--54)	67 (63-71)	0.01	<0.0001
MRC dyspnoea score	3.2 (2.7-3.7)	2.4 (1.9-2.9)	1.6 (1.0-2.1)	n.s.	0.05
Exacerbation frequency (previous year)	1.7 (0.1-3.2)	3.4 (2.6-4.1)	2.5 (1.7-3.2)	n.s.	n.s.
ICS use (previous year)	14 (43.8%)	18 (64.3%)	17 (54.8%)	n.s.	n.s.
Long term Oral antibiotics use (previous year)	2 (6.3%)	9 (32.1%)	11 (35.4%)	0.01	n.s.
Inhaled antibiotics use (previous year)	0 (0%)	1 (3.6%)	6 (19.4%)	n.s.	n.s.

Figure 1. Relative abundance of lung microbiota at the phylum level (A) and Shannon diversity index (B) in COPD, bronchiectasis and the 'COPD-bronchiectasis association'.

* p<0.05; ** p<0.01.

Figure 2. PCA of patients with COPD, bronchiectasis, and the 'COPD-bronchiectasis association' based on sputum proteome and microbiome profiles

A. The scores plot of PCA. Each dot represents a patient. R2X[1]=0.08, R2X[2]=0.04.. Black eclipse: Hostelling T2 95% confidence.
B. The loadings plot of PCA showing the key protein and bacterial species driving the largest variability within the dataset observed in the scores plot (A). LCP1: Lymphocyte cytosolic protein 1, CAT: catalase, MMP-9: metalloproteinase 9, CAMP: cathelicidin antimicrobial peptide, PIP: prolactin induced protein, TIMP1: tissue inhibitor of metalloproteinases 1, PIGR: polymeric immunoglobulin receptor, and IgK@: immunoglobulin kappa chain, ACTG1: actin G1, EZR: ezrin, GSN: gelsolin, AHSG: alpha 2-HS glycoprotein, EPX: eosinophil peroxidase, CLC: galectin-10. Key protein and microbiome variables for principal component 1 are shown in red and blue, respectively, and those for principal component 2 are shown in green.

Figure 3. PLS-DA of patients with COPD and the 'COPD-bronchiectasis association' based on combined sputum proteome and microbiome profiles.

A. The scores plot of PLS-DA. R2X[1]=0.08, R2X[2]=0.04. Q2 [cum]= 0.50 (range = 0.42-0.58 following six times of the random permutation procedure previously described.(26) $p = 3.5 \times 10^{-9}$ (CV-ANOVA).
B. The loadings plot of PLS-DA showing the key protein and bacterial species driving the differences observed in the scores plot (A). ORM1: orosomucoid 1, BPIFB1: bactericidal/permeability-increasing protein-like 1 (BPIFB1), SLPI: secretory leukocyte peptidase Inhibitor: PIGR: polymeric immunoglobulin receptor; IgJ: immunoglobulin J chain. Green eclipse: key variables whose expression/abundance are positively associated with COPD. Red eclipse: key variables whose expression/abundance are positively associated with the COPD-bronchiectasis association. Key protein and microbiome features driving the separation in A are

shown in red and blue, respectively. Those labelled in black are the variables drive the within-group variability.

Figure 4. PCA of patients with COPD, bronchiectasis, and the 'COPD-bronchiectasis association' based on sputum proteins and bacterial culture in the validation cohort

Validation of key findings in the first cohort using targeted analyses in an independent cohort of 32 patients with COPD, 31 patients with bronchiectasis and 28 patients with bronchiectasis-COPD association.

A. The scores plot of PCA on targeted sputum proteins and bacteria in patients with COPD, bronchiectasis and the 'COPD-bronchiectasis association'. R2X[1] = 0.15, R2X[2] =0.12.

B. The loadings plot of PCA showing the key proteins and bacterial species driving the differences observed in the scores plot (D). IL-4: interleukin-4, IL-5: interleukin-5, IL-8: interleukin-8, GM-CSF: granulocyte-macrophage colony-stimulating factor, TSLP: thymic stromal lymphopoietin, TARC: thymus and activation-regulated chemokine, DNA: sputum DNA content, NE: neutrophil elastase.

Figure 5. Five endotypes based on proteome and microbiome profiles of patients with COPD, bronchiectasis and the 'COPD-bronchiectasis association'.

Key features observed in sputum proteome and microbiome are plotted in a 3 dimensional space with neutrophilic inflammation in x axis, protease inhibitors/immunoglobulins/ MMP8/MMP9 in y axis, and eosinophilic inflammation in z axis. See Results and Discussion sections for further details. The eclipses represent hypothetical boundaries of three disease groups.

References

1. Traversi L, Miravitlles M, Martinez-Garcia M, Shteinberg M, Bossios A, Dimakou K, Jacob J, Hurst JR, Paggiaro PL, Ferri F, Hillas G, Vogel-Claussen J, Dettmer S, Aliberti S, Chalmers JD, E. P. ROSE: Radiology, Obstruction, Symptoms and Exposure, a Delphi consensus definition of the association of COPD and Bronchiectasis by the EMBARC Airways Working Group. *ERJ Open Res* 2021; (in press).
2. Polverino E, Dimakou K, Hurst J, Martinez-Garcia MA, Miravitlles M, Paggiaro P, Shteinberg M, Aliberti S, Chalmers JD. The overlap between bronchiectasis and chronic airway diseases: state of the art and future directions. *Eur Respir J* 2018; 52.
3. Aksamit TR, O'Donnell AE, Barker A, Olivier KN, Winthrop KL, Daniels MLA, Johnson M, Eden E, Griffith D, Knowles M, Metersky M, Salathe M, Thomashow B, Tino G, Turino G, Carretta B, Daley CL, Bronchiectasis Research Registry C. Adult Patients With Bronchiectasis: A First Look at the US Bronchiectasis Research Registry. *Chest* 2017; 151: 982-992.
4. Chalmers JD, Aliberti S, Polverino E, Vendrell M, Crichton M, Loebinger M, Dimakou K, Clifton I, van der Eerden M, Rohde G, Murris-Espin M, Masefield S, Gerada E, Shteinberg M, Ringshausen F, Haworth C, Boersma W, Rademacher J, Hill AT, Aksamit T, O'Donnell A, Morgan L, Milenkovic B, Tramma L, Neves J, Menendez R, Paggiaro P, Botnaru V, Skrgat S, Wilson R, Goeminne P, De Soyza A, Welte T, Torres A, Elborn JS, Blasi F. The EMBARC European Bronchiectasis Registry: protocol for an international observational study. *ERJ Open Res* 2016; 2.
5. Hurst JR, Elborn JS, De Soyza A, Consortium B-U. COPD-bronchiectasis overlap syndrome. *Eur Respir J* 2015; 45: 310-313.
6. Molino A, Vitale C, Valente T, D'Amato M, Mormile M, Imitazione P, Maglio A, Vatrella A. Impact of Bronchiectasis on COPD Exacerbations. *Int J of Thorax* 2018; 1: 1-7.
7. Labaki WW, Han MK. Impact of bronchiectasis on the frequency and severity of respiratory exacerbations in COPD. *Int J Chron Obstruct Pulmon Dis* 2018; 13: 2335-2338.
8. Pitassi M, Vitale C, D'Amato M, Stanziola A, Mormile M, Molino A. Impact of bronchiectasis on duration of hospitalization in patients with exacerbations of COPD. *European Respiratory Journal* 2015; 46.
9. Martinez-Garcia MA, de la Rosa Carrillo D, Soler-Cataluna JJ, Donat-Sanz Y, Serra PC, Lerma MA, Ballestin J, Sanchez IV, Selma Ferrer MJ, Dalfo AR, Valdecillos MB. Prognostic value of bronchiectasis in patients with moderate-to-severe chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013; 187: 823-831.
10. Du Q, Jin J, Liu X, Sun Y. Bronchiectasis as a Comorbidity of Chronic Obstructive Pulmonary Disease: A Systematic Review and Meta-Analysis. *PLoS One* 2016; 11: e0150532.
11. Martinez-Garcia MA, Polverino E, Aksamit T. Bronchiectasis and Chronic Airway Disease: It Is Not Just About Asthma and COPD. *Chest* 2018; 154: 737-739.
12. Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet* 2008; 372: 1107-1119.
13. Agusti A, Bel E, Thomas M, Vogelmeier C, Brusselle G, Holgate S, Humbert M, Jones P, Gibson PG, Vestbo J, Beasley R, Pavord ID. Treatable traits: toward precision medicine of chronic airway diseases. *Eur Respir J* 2016; 47: 410-419.
14. McDonald VM, Fingleton J, Agusti A, Hiles SA, Clark VL, Holland AE, Marks GB, Bardin PP, Beasley R, Pavord ID, Wark PAB, Gibson PG, participants of the Treatable Traits

Down Under International W, Treatable Traits Down Under International Workshop p. Treatable traits: a new paradigm for 21st century management of chronic airway diseases: Treatable Traits Down Under International Workshop report. *Eur Respir J* 2019; 53.
15. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol* 2017; 18: 83.
16. Dicker AJ, Huang JTJ, Lonergan M, Keir HR, Fong CJ, Tan B, Cassidy AJ, Finch S, Mullerova H, Miller BE, Tal-Singer R, Chalmers JD. The sputum microbiome, airway inflammation, and mortality in chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 2021; 147: 158-167.
17. Keir HR, Dicker A, Lonergan M, Crichton M, Miller BE, Tal-Singer R, Chalmers JD. Clinical endotypes of exacerbation are associated with differences in microbial composition and diversity in COPD. *Eur Respir J* 2020; 56.
18. Chalmers JD, Moffitt KL, Suarez-Cuartin G, Sibila O, Finch S, Furrie E, Dicker A, Wrobel K, Elborn JS, Walker B, Martin SL, Marshall SE, Huang JT, Fardon TC. Neutrophil Elastase Activity is Associated with Exacerbations and Lung Function Decline in Bronchiectasis. *Am J Respir Crit Care Med* 2017; 195: 1384-1393.
19. Short PM, Lipworth SI, Elder DH, Schembri S, Lipworth BJ. Effect of beta blockers in treatment of chronic obstructive pulmonary disease: a retrospective cohort study. *BMJ* 2011; 342: d2549.
20. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; 246: 697-722.
21. Chalmers JD, Goeminne P, Aliberti S, McDonnell MJ, Lonni S, Davidson J, Poppelwell L, Salih W, Pesci A, Dupont LJ, Fardon TC, De Soyza A, Hill AT. The bronchiectasis severity index. An international derivation and validation study. *Am J Respir Crit Care Med* 2014; 189: 576-585.
22. Singh D, Agusti A, Anzueto A, Barnes PJ, Bourbeau J, Celli BR, Criner GJ, Frith P, Halpin DMG, Han M, Lopez Varela MV, Martinez F, Montes de Oca M, Papi A, Pavord ID, Roche N, Sin DD, Stockley R, Vestbo J, Wedzicha JA, Vogelmeier C. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2019. *Eur Respir J* 2019; 53.
23. Quittner AL, O'Donnell AE, Salathe MA, Lewis SA, Li X, Montgomery AB, O'Riordan TG, Barker AF. Quality of Life Questionnaire-Bronchiectasis: final psychometric analyses and determination of minimal important difference scores. *Thorax* 2015; 70: 12-20.
24. Reiff DB, Wells AU, Carr DH, Cole PJ, Hansell DM. CT findings in bronchiectasis: limited value in distinguishing between idiopathic and specific types. *AJR Am J Roentgenol* 1995; 165: 261-267.
25. Worley B, Powers R. PCA as a practical indicator of OPLS-DA model reliability. *Curr Metabolomics* 2016; 4: 97-103.
26. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, Rutledge DN, Savarin P. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Mol Biosyst* 2015; 11: 13-19.
27. Eriksson L, Trygg J, Wold S. CV-ANOVA for significance testing of PLS and OPLS® models. *Journal of Chemometrics* 2008; 22: 594-600.
28. Candiano G, Bruschi M, Pedemonte N, Caci E, Liberatori S, Bini L, Pellegrini C, Vigano M, O'Connor BJ, Lee TH, Galietta LJ, Zegarra-Moran O. Gelsolin secretion in

interleukin-4-treated bronchial epithelia and in asthmatic airways. *Am J Respir Crit Care Med* 2005; 172: 1090-1096.
29. Dicker AJ, Crichton ML, Pumphrey EG, Cassidy AJ, Suarez-Cuartin G, Sibila O, Furrie E, Fong CJ, Ibrahim W, Brady G, Einarsson GG, Elborn JS, Schembri S, Marshall SE, Palmer CNA, Chalmers JD. Neutrophil extracellular traps are associated with disease severity and microbiota diversity in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 2018; 141: 117-127.
30. Chalmers JD, Boersma W, Lonergan M, Jayaram L, Crichton ML, Karalus N, Taylor SL, Martin ML, Burr LD, Wong C, Altenburg J. Long-term macrolide antibiotics for the treatment of bronchiectasis in adults: an individual participant data meta-analysis. *Lancet Respir Med* 2019; 7: 845-854.
31. Poachanukoon O, Koontongkaew S, Monthanapisut P, Pattanacharoenchai N. Macrolides attenuate phorbol ester-induced tumor necrosis factor-alpha and mucin production from human airway epithelial cells. *Pharmacology* 2014; 93: 92-99.
32. Keir HR, Shoemark A, Dicker AJ, Perea L, Pollock J, Giam YH, Suarez-Cuartin G, Crichton ML, Lonergan M, Oriano M, Cant E, Einarsson GG, Furrie E, Elborn JS, Fong CJ, Finch S, Rogers GB, Blasi F, Sibila O, Aliberti S, Simpson JL, Huang JTJ, Chalmers JD. Neutrophil extracellular traps, disease severity, and antibiotic response in bronchiectasis: an international, observational, multicohort study. *Lancet Respir Med* 2021.
33. Chalmers JD, Haworth CS, Metersky ML, Loebinger MR, Blasi F, Sibila O, O'Donnell AE, Sullivan EJ, Mange KC, Fernandez C, Zou J, Daley CL, Investigators W. Phase 2 Trial of the DPP-1 Inhibitor Brensocatib in Bronchiectasis. *N Engl J Med* 2020; 383: 2127-2137.
34. Eklof J, Sorensen R, Ingebrigtsen TS, Sivapalan P, Achir I, Boel JB, Bangsborg J, Ostergaard C, Dessau RB, Jensen US, Browatzki A, Lapperre TS, Janner J, Weinreich UM, Armbruster K, Wilcke T, Seersholm N, Jensen JUS. Pseudomonas aeruginosa and risk of death and exacerbations in patients with chronic obstructive pulmonary disease: an observational cohort study of 22 053 patients. *Clin Microbiol Infect* 2020; 26: 227-234.
35. Iyer Parameswaran G, Murphy TF. Chronic obstructive pulmonary disease: role of bacteria and updated guide to antibacterial selection in the older patient. *Drugs Aging* 2009; 26: 985-995.
36. Altenburg J, de Graaff CS, Stienstra Y, Sloos JH, van Haren EH, Koppers RJ, van der Werf TS, Boersma WG. Effect of azithromycin maintenance treatment on infectious exacerbations among patients with non-cystic fibrosis bronchiectasis: the BAT randomized controlled trial. *JAMA* 2013; 309: 1251-1259.
37. Wilson CB, Jones PW, O'Leary CJ, Hansell DM, Cole PJ, Wilson R. Effect of sputum bacteriology on the quality of life of patients with bronchiectasis. *Eur Respir J* 1997; 10: 1754-1760.
38. Evans SA, Turner SM, Bosch BJ, Hardy CC, Woodhead MA. Lung function in bronchiectasis: the influence of Pseudomonas aeruginosa. *Eur Respir J* 1996; 9: 1601-1604.
39. Rogers GB, Zain NM, Bruce KD, Burr LD, Chen AC, Rivett DW, McGuckin MA, Serisier DJ. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Ann Am Thorac Soc* 2014; 11: 496-503.

40. Zhu A, Ge D, Zhang J, Teng Y, Yuan C, Huang M, Adcock IM, Barnes PJ, Yao X. Sputum myeloperoxidase in chronic obstructive pulmonary disease. *Eur J Med Res* 2014; 19: 12.
41. Kesimer M, Ford AA, Ceppe A, Radicioni G, Cao R, Davis CW, Doerschuk CM, Alexis NE, Anderson WH, Henderson AG, Barr RG, Bleecker ER, Christenson SA, Cooper CB, Han MK, Hansel NN, Hastie AT, Hoffman EA, Kanner RE, Martinez F, Paine R, 3rd, Woodruff PG, O'Neal WK, Boucher RC. Airway Mucin Concentration as a Marker of Chronic Bronchitis. *N Engl J Med* 2017; 377: 911-922.
42. Gompertz S, Bayley DL, Hill SL, Stockley RA. Relationship between airway inflammation and the frequency of exacerbations in patients with smoking related COPD. *Thorax* 2001; 56: 36-41.
43. Yang YM, Guo YF, Zhang HS, Sun TY. Antimicrobial peptide LL-37 circulating levels in chronic obstructive pulmonary disease patients with high risk of frequent exacerbations. *J Thorac Dis* 2015; 7: 740-745.
44. Sibila O, Perea L, Canto E, Shoemark A, Cassidy D, Smith AH, Suarez-Cuartin G, Rodrigo-Troyano A, Keir HR, Oriano M, Ong S, Vidal S, Blasi F, Aliberti S, Chalmers JD. Antimicrobial peptides, disease severity and exacerbations in bronchiectasis. *Thorax* 2019; 74: 835-842.
45. Ridley C, Thornton DJ. Mucins: the frontline defence of the lung. *Biochem Soc Trans* 2018; 46: 1099-1106.
46. Livraghi-Butrico A, Grubb BR, Wilkinson KJ, Volmer AS, Burns KA, Evans CM, O'Neal WK, Boucher RC. Contribution of mucus concentration and secreted mucins Muc5ac and Muc5b to the pathogenesis of muco-obstructive lung disease. *Mucosal Immunol* 2017; 10: 395-407.
47. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, Alexander SN, Bellinghausen LK, Song AS, Petrova YM, Tuvim MJ, Adachi R, Romo I, Bordt AS, Bowden MG, Sisson JH, Woodruff PG, Thornton DJ, Rousseau K, De la Garza MM, Moghaddam SJ, Karmouty-Quintana H, Blackburn MR, Drouin SM, Davis CW, Terrell KA, Grubb BR, O'Neal WK, Flores SC, Cota-Gomez A, Lozupone CA, Donnelly JM, Watson AM, Hennessy CE, Keith RC, Yang IV, Barthel L, Henson PM, Janssen WJ, Schwartz DA, Boucher RC, Dickey BF, Evans CM. Muc5b is required for airway defence. *Nature* 2014; 505: 412-416.
48. Ehre C, Worthington EN, Liesman RM, Grubb BR, Barbier D, O'Neal WK, Sallenave JM, Pickles RJ, Boucher RC. Overexpressing mouse model demonstrates the protective role of Muc5ac in the lungs. *Proc Natl Acad Sci U S A* 2012; 109: 16528-16533.
49. Evans CM, Raclawska DS, Ttofali F, Liptzin DR, Fletcher AA, Harper DN, McGing MA, McElwee MM, Williams OW, Sanchez E, Roy MG, Kindrachuk KN, Wynn TA, Eltzschig HK, Blackburn MR, Tuvim MJ, Janssen WJ, Schwartz DA, Dickey BF. The polymeric mucin Muc5ac is required for allergic airway hyperreactivity. *Nat Commun* 2015; 6: 6281.
50. Koeppen M, McNamee EN, Brodsky KS, Aherne CM, Faigle M, Downey GP, Colgan SP, Evans CM, Schwartz DA, Eltzschig HK. Detrimental role of the airway mucin Muc5ac during ventilator-induced lung injury. *Mucosal Immunol* 2013; 6: 762-775.
51. Radicioni G, Ceppe A, Ford AA, Alexis NE, Barr RG, Bleecker ER, Christenson SA, Cooper CB, Han MK, Hansel NN, Hastie AT, Hoffman EA, Kanner RE, Martinez FJ, Ozkan E, Paine R, 3rd, Woodruff PG, O'Neal WK, Boucher RC, Kesimer M. Airway mucin MUC5AC and MUC5B concentrations and the initiation and progression of chronic

obstructive pulmonary disease: an analysis of the SPIROMICS cohort. *Lancet Respir Med* 2021; 9: 1241-1254.

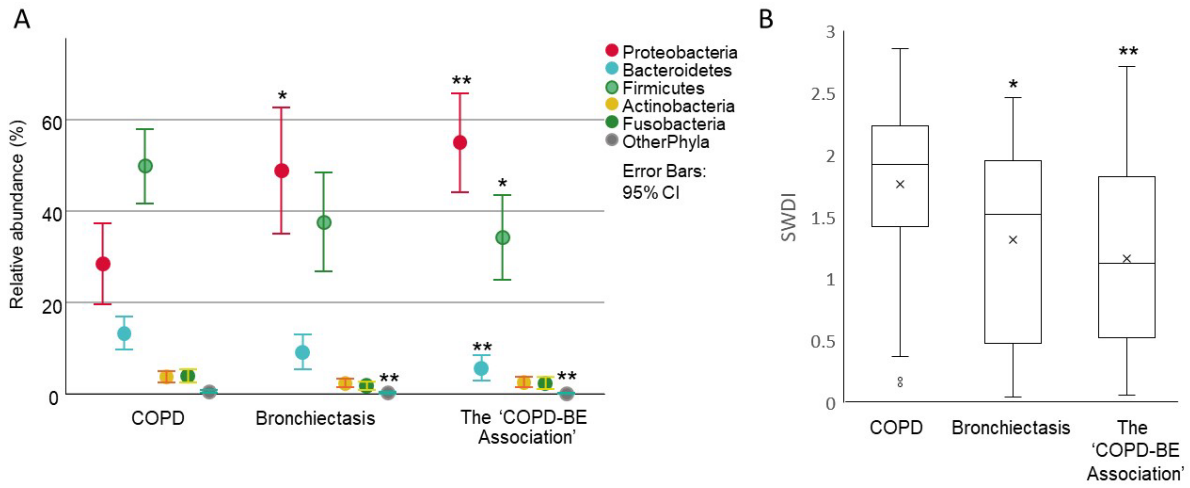

Figure 1

Figure 1. Relative abundance of lung microbiota at the phylum level (A) and Shannon diversity index (B) in COPD, bronchiectasis and the 'COPD-bronchiectasis association'.

* p<0.05; ** p<0.01 vs COPD.

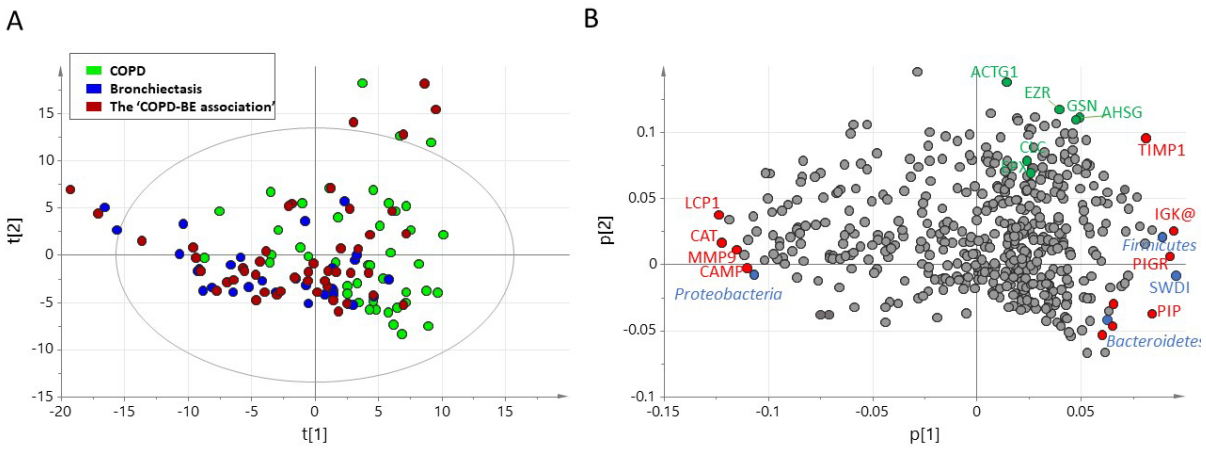

Figure 2

Figure 2. PCA of patients with COPD, bronchiectasis, and the 'COPD-bronchiectasis association' based on sputum proteome and microbiome profiles

A. The scores plot of PCA. Each dot represents a patient. R2X[1]=0.08, R2X[2]=0.04. Black eclipse: Hostelling T2 95% confidence.

B. The loadings plot of PCA showing the key protein and bacterial species driving the largest variability within the dataset observed in the scores plot (A). LCP1: Lymphocyte cytosolic protein 1, CAT: catalase, MMP-9: metalloproteinase 9, CAMP: cathelicidin antimicrobial peptide, PIP: prolactin induced protein, TIMP1: tissue inhibitor of metalloproteinases 1, PIGR: polymeric immunoglobulin receptor, and IgK@: immunoglobulin kappa chain, ACTG1: actin G1, EZR: ezrin, GSN: gelsolin, AHSG: alpha 2-HS glycoprotein, EPX: eosinophil peroxidase, CLC: galectin-10. Key protein and microbiome variables for principal component 1 are shown in red and blue, respectively, and those for principal component 2 are shown in green.

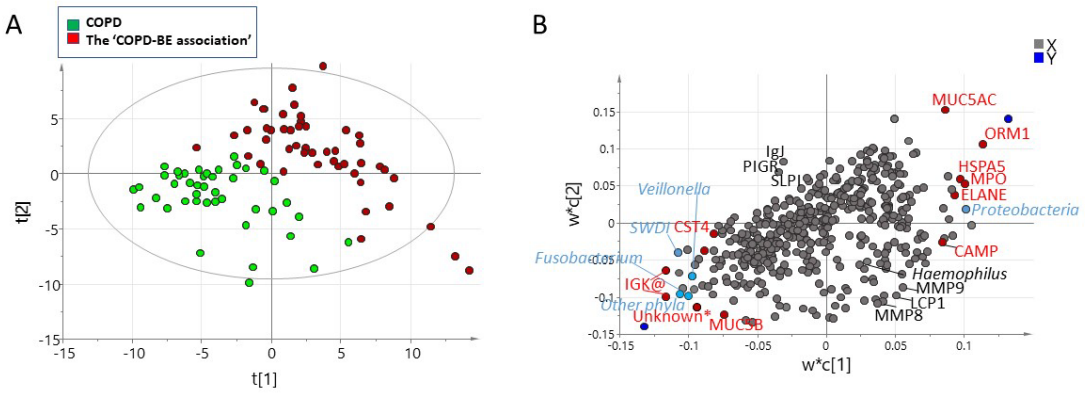

Figure 3

Figure 3. PLS-DA of patients with COPD and the 'COPD-bronchiectasis association' based on combined sputum proteome and microbiome profiles.

A. The scores plot of PLS-DA. R2X[1]=0.08, R2X[2]=0.04. Q2 [cum]= 0.50 (range = 0.42-0.58 following six times of the random permutation procedure previously described.(26) p = 3.5 x 10^{-9} (CV-ANOVA).

B. The loadings plot of PLS-DA showing the key protein and bacterial species driving the differences observed in the scores plot (A). ORM1: orosomucoid 1, BPIFB1: bactericidal/permeability-increasing protein-like 1 (BPIFB1), SLPI: secretory leukocyte peptidase Inhibitor: PIGR: polymeric immunoglobulin receptor; IgJ: immunoglobulin J chain. Green eclipse: key variables whose expression/abundance are positively associated with COPD. Red eclipse: key variables whose expression/abundance are positively associated with the COPD-bronchiectasis association. Key protein and microbiome features driving the separation in A are shown in red and blue, respectively. Those labelled in black are the variables drive the within-group variability.

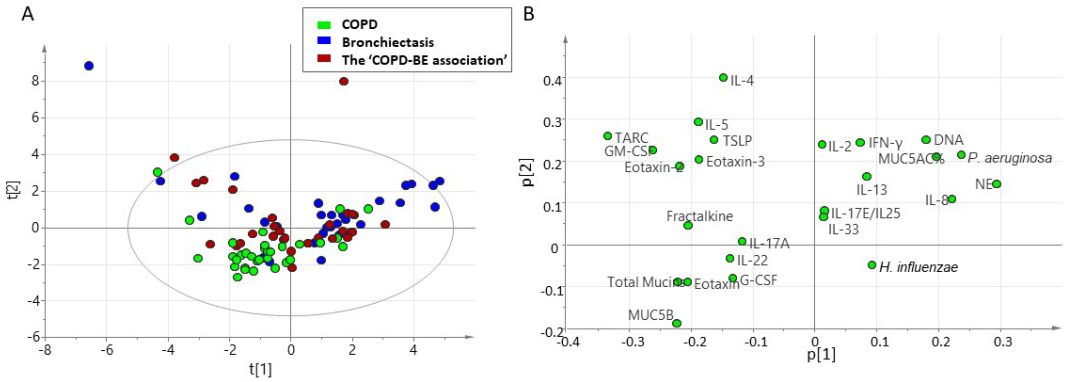

Figure 4

Figure 4. PCA of patients with COPD, bronchiectasis, and the 'COPD-bronchiectasis association' based on sputum proteins and bacterial culture in the validation cohort

Validation of key findings in the first cohort using targeted analyses in an independent cohort of 32 patients with COPD, 31 patients with bronchiectasis and 28 patients with bronchiectasis-COPD association.

A. The scores plot of PCA on targeted sputum proteins and bacteria in patients with COPD, bronchiectasis and the 'COPD-bronchiectasis association'. R2X[1] = 0.15, R2X[2] =0.12.

B. The loadings plot of PCA showing the key proteins and bacterial species driving the differences observed in the scores plot (D). IL-4: interleukin-4, IL-5: interleukin-5, IL-8: interleukin-8, GM-CSF: granulocyte-macrophage colony-stimulating factor, TSLP: thymic stromal lymphopoietin, TARC: thymus and activation-regulated chemokine, DNA: sputum DNA content, NE: neutrophil elastase.

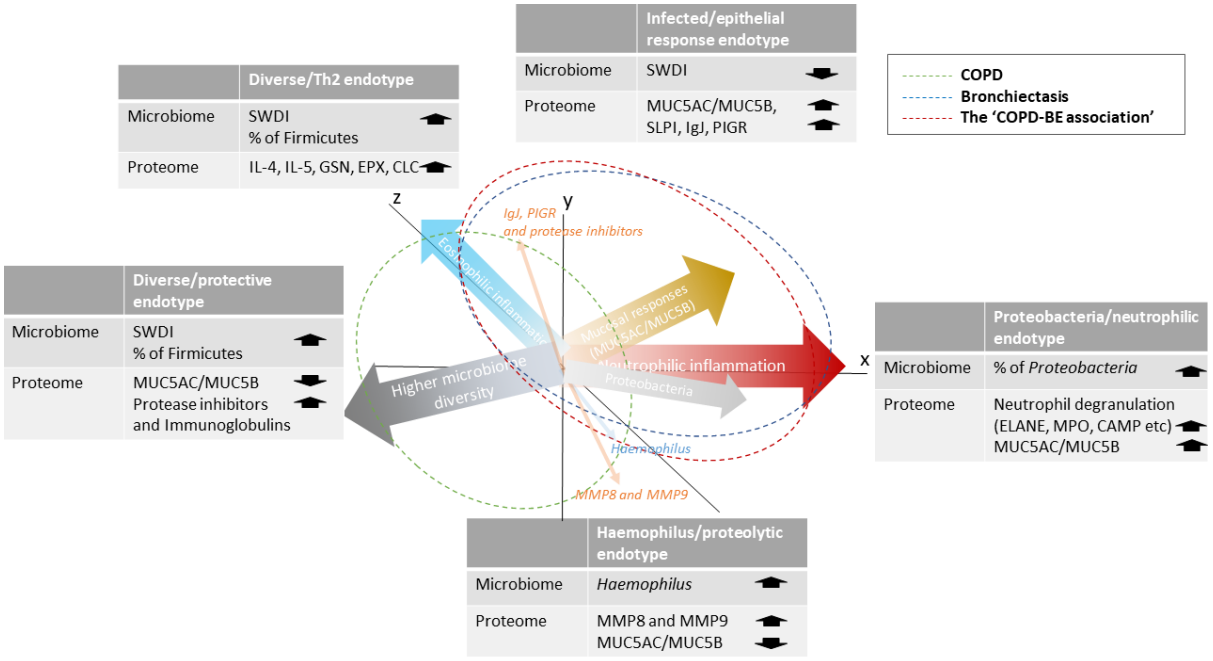

Figure 5

Figure 5. Five endotypes based on proteome and microbiome profiles of patients with COPD, bronchiectasis and the 'COPD-bronchiectasis association'.

Key features observed in sputum proteome and microbiome are plotted in a 3 dimensional space with neutrophilic inflammation in x axis, protease inhibitors/immunoglobulins/ MMP8/MMP9 in y axis, and eosinophilic inflammation in z axis. See Results and Discussion sections for further details. The eclipses represent hypothetical boundaries of three disease groups.