Sputum Proteomics in non-Tuberculous Mycobacterial Lung Disease

Hull, Rebecca C.; Huang, Jeffrey T. J.; Barton, Alun K.; Keir, Holly R.; Ellis, Huw; Cookson, William

Published in:
Chest

DOI:
10.1016/j.chest.2021.11.014

Publication date:
2022

Citation for published version (APA):
Title: Sputum Proteomics in non-Tuberculous Mycobacterial Lung Disease

Authors: Rebecca C Hull, BSc1,2, Jeffrey TJ Huang PhD3, Alun K Barton3, Holly R Keir BSc1, Huw Ellis PhD5,6, William Cookson PhD5,6, Miriam Moffatt PhD5,6, Michael Loebinger PhD5,6*, James D Chalmers PhD4*

*contributed equally

1. Department of Infection, Immunity and Cardiovascular Diseases, University of Sheffield, Beech Hill Road, Sheffield, UK
2. Florey Institute, University of Sheffield, Beech Hill Road, Sheffield, UK.
3. Division of Systems Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, Scotland, UK.
4. Division of Molecular and Clinical Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, Scotland, UK.
5. Royal Brompton and Harefield NHS Foundation Trust, London, UK
6. National Heart and Lung Institute, Imperial College, London, UK.

Corresponding author: James D. Chalmers, Division of Molecular and Clinical Medicine, University of Dundee, Ninewells Hospital and Medical School, Dundee, DD1 9SY. jchalmers@dundee.ac.uk

Summary conflict of interest statements: RCH, JTJH, AKB, HRK, HE, WC and MM have no conflicts of interest to disclosure. ML declares consultancy and research grants from Insmed, JDC reports consultancy of research grants from AstraZeneca, Boehringer-Ingelheim, Chiesi, Gilead Sciences, Grifols, Insmed, Janssen, Novartis and Zambon

Funding: 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 GSK/British Lung Foundation Chair of Respiratory Research and a Scottish Senior Fellowship from the Chief Scientist Office. RCH is supported by a scholarship from the Medical Research Foundation National PhD Training Programme in Antimicrobial Resistance Research (MRF-145-0004-TPG-AVISO) and the AMR cross-council funding from the MRC to the SHIELD consortium “Optimising Innate Host Defence to Combat Antimicrobial Resistance” MRNO2995X/1.

An abstract of part of this work was presented online as an e-poster in the ERS congress 2021, 5-8/9/2021
**Keywords**: biomarkers, bronchiectasis, COPD, cystic fibrosis, NTM

**Abbreviations:**

BE- Bronchiectasis  
CF- Cystic fibrosis  
COPD- Chronic obstructive pulmonary disease  
FEV1- Forced expiratory volume in 1 second  
FVC- Forced vital capacity  
PCA- Principal component analysis  
MAC- *Mycobacterium avium* complex  
NET- Neutrophil extracellular trap  
NTM- Non-tuberculous mycobacterial
Abstract

Background: Non-tuberculous mycobacterial (NTM) infections are difficult to diagnose and treat. Biomarkers to identify patients with active infection or at risk of disease progression would have clinical utility. Sputum is the most frequently used matrix for diagnosis of NTM lung disease.

Research Question: Can sputum proteomics be used to identify NTM associated inflammatory profiles in sputum?

Study Design and Methods: Patients with NTM lung disease and a matched cohort of patients with chronic obstructive pulmonary disease (COPD), bronchiectasis (BE) and cystic fibrosis (CF) without NTM lung disease were enrolled from two hospitals in the UK. Liquid chromatography-tandem mass spectrometry was used to identify proteomic biomarkers associated with underlying diagnosis (COPD, BE, CF), the presence of NTM lung disease defined by ATS/IDSA criteria and severity of NTM. A subset of patients receiving guideline concordant NTM treatment were studied to identify protein changes associated with treatment response.

Results: We analysed 95 sputum samples from 55 subjects (21 BE, 19 COPD, 15 CF). Underlying disease and infection with Pseudomonas aeruginosa were the strongest drivers of sputum protein profiles. Comparing protein abundance in COPD, BE and CF showed 12 proteins were upregulated in CF including MPO, AZU1, CTSG, CAT and RNASE3 with 21 proteins downregulated including SCGB1A1, IGFBP2, SFTPB, GC and CFD. Across CF, BE and COPD, NTM infection (n=15) was not associated with statistically significant differences in sputum protein profiles compared to those without NTM. 2 proteins associated with iron chelation were significantly downregulated in severe NTM disease. NTM treatment was associated with heterogeneous changes in sputum protein profile. NTM patients with a decrease in immune response proteins had a subjective symptomatic improvement.

Interpretation: Sputum proteomics identified candidate biomarkers of NTM severity and treatment response, however underlying lung disease and typical bacterial pathogens such as P. aeruginosa are also key determinants of sputum proteome profile.
Non-tuberculous Mycobacterial (NTM) lung infections are an increasing clinical problem worldwide.\(^1\)\(^2\) Patients experience a clinical syndrome of progressive respiratory symptoms, lung damage and increased mortality.\(^3\)\(^4\)\(^5\)\(^6\)

NTM refers to members of the genus Mycobacterium excluding those that form part of the \(M.\) \(tuberculosis\) complex.\(^7\) They are environmental microorganisms and opportunistic pathogens of humans, primarily affecting patients with underlying structural lung disease or immunosuppression. The most frequent NTM causing human lung infections are \(Mycobacterium\) \(avium\) complex (MAC), \(Mycobacterium\) \(abscessus\) complex, \(M.\) \(kansasii\), \(M.\) \(simiae\), \(M.\) \(malmoense\), \(M.\) \(chelae\), \(M.\) \(fortuitum\) and \(M.\) \(szulgai\) but over 180 species have now been described.\(^8\)\(^9\)

NTM lung infections are extremely challenging. Diagnosis is often delayed due to insidious presentation and difficulties in differentiating between transient colonization of the airways and persistent infection.\(^10\) The decision to treat is challenging, because some patients experience progressive lung damage due to NTM while others do not. Success of treatment has been estimated at an average of 60% for MAC lung disease\(^11\) and 45.6% for \(M.\) \(abscessus\) lung disease.\(^12\)\(^13\) Treatment is prolonged, typically 12 months after sputum culture conversion has been achieved, and is toxic, requiring combinations of multiple antibiotics dependent on the infecting species and drug susceptibilities.\(^2\)\(^14\)

NTM lung disease therefore represents a clinical entity where biomarkers that could guide clinical decision making would have a high degree of clinical utility. Biomarkers indicating the presence of NTM could identify infections at an early stage, since culture can take several weeks, while biomarkers indicating the severity of lung disease or risk of progression could aid decision making around treatment. Biomarkers that indicate treatment response could support decisions around duration of treatment. Bronchiectasis is the most common underlying condition in patients with NTM lung disease\(^15\) and in a recent study we demonstrated that sputum proteomics is a powerful tool for biomarker discovery in patients with bronchiectasis.\(^16\) Sputum proteomics in 40 patients with bronchiectasis (20 severe vs 20 mild) identified 96 proteins that were associated with severity of disease, while a study of patients with acute exacerbations (\(N=20\)) found 39 proteins significantly associated with antibiotic response.\(^16\)
In this study we utilised sputum proteomics in a cohort of patients with and without NTM Lung disease to establish the potential for this technique to identify biomarkers for use in NTM lung disease.

Outcomes and objectives
Based on our prior work\textsuperscript{16,20} showing a high degree of diversity in the bronchiectasis and chronic obstructive pulmonary disease (COPD) sputum protein profiles we initially compared COPD, bronchiectasis and cystic fibrosis (CF), irrespective of NTM status to determine the contribution of underlying disease to variation in the sputum proteome. We then compared those with NTM vs those without NTM infection and finally the third objective was to examine proteomic changes over time in patients on treatment.
Methods

Study design
Patients with NTM lung disease and control subjects with underlying lung conditions without NTM infection were recruited from two specialist respiratory centres at Ninewells Hospital, Dundee, UK and Royal Brompton Hospital, London, UK. The study was approved by the East of Scotland Research Ethics Committee and all patients gave written informed consent to participate.

Patients and samples
NTM lung disease was defined according to consensus guidelines requiring microbiological, clinical and radiological confirmation of disease. All patients had a positive sputum sample within 3 months prior to enrolment. Severe NTM infection was defined by the presence of cavitary disease or severe symptoms as judged by the treating clinician. Bronchiectasis was defined by the presence of irreversible bronchial dilation on high resolution CT scan and the presence of respiratory symptoms consistent with the disease.

Chronic Obstructive Pulmonary Disease was diagnosed based on the GOLD definition along with evidence of post-bronchodilator airflow obstruction reflected by a forced expiratory volume in 1 second (FEV1) / forced vital capacity (FVC) ratio less than 0.7. Cystic fibrosis was defined by the presence of two disease causing CFTR mutations.

For inclusion patients had to meet one of the above disease definitions and also meet the following inclusion criteria: age>18 years, clinically stable and free from antibiotic treatment for at least 4 weeks (excluding prophylactic antibiotic treatment such as long term macrolides), able to provide a spontaneous sputum sample at the study visit and able to give informed consent. Patients without NTM lung disease were required to have 3 sputum samples sent for NTM culture over a 3 month period and for each of these to be negative prior to inclusion in the study. Patients with NTM lung disease were enrolled prior to the commencement of antibiotic treatment with the first sputum sample included in the analysis being prior to antibiotic therapy in all cases.

Patients were enrolled consecutively with the aim to have 15 patients with NTM disease and at least 2 disease controls for each NTM patient. No attempt was made to match for age, sex or BMI as prior studies in both COPD and bronchiectasis found no relationships between sputum proteome and these variables.

11 NTM patients subsequently had sputum samples obtained at 4 further time points – week 1-2, month 2-4, month 6-9 and month 10-15. Sputum was sent for AAFB staining and culture to
determine the date of sputum culture conversion and allow analysis of changes in protein profiles associated with treatment response. Clinical improvement was defined by the treating clinician as an improvement in symptoms and quality of life with or without accompanying improvements in radiology. Lack of clinical improvement was determined where patients subjectively did not report improved symptoms or quality of life from baseline. Bacterial co-infection at baseline was determined by standard bacterial culture of sputum samples.

**Sputum proteomics**
Sputum protein profiling was performed using a label-free shotgun proteomic workflow with a nanoflow liquid chromatography system (Agilent 1200, Agilent) linking to an Orbitrap mass spectrometer (LTQ-Orbitrap, Thermo Scientific). Protein identification and label-free quantification were carried out using Maxquant (version 1.4.1.2) against Uniprot-human database (version 2019-09-29). FDR for protein identification was set to 1% at peptide spectrum matches level.

**Statistical analysis**
All statistical analysis was performed using R (version 4.0.3). Dimensional reduction of 693 protein abundances per patient was carried out by principal component analysis (PCA) (prcomp in default “stats” package). Visualising PCAs was carried out through ggbiplot (R package ggbiplot) with ellipses around groupings represent 68% CI. Relative abundance of proteins between two disease groups were calculated using t-test with Benjamini & Hochberg correction for multiple comparisons (R package rstatix, stats). Linear regression was performed with lm model (default “stats” package). Linear mixed effects model (R package lme4) of protein abundances over time was used on data with patients ID as a random effect followed by a 2-way ANOVA (anova in default “stats” package) to determine interactions.
Results

The study included 55 participants. 21 patients with bronchiectasis, 19 patients with COPD and 15 patients with CF. 15 patients had NTM infection of which 11 had repeat visits over 10-15 months. Of these the following species were identified (*M. avium, M. abscessus, M. xenopi*).

Characteristics of the patients are shown in table 1.

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Bronchiectasis</th>
<th>COPD</th>
<th>Cystic Fibrosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-NTM</td>
<td>NTM</td>
<td>Non-NTM</td>
<td>NTM</td>
</tr>
<tr>
<td>Age</td>
<td>73.6 (9.25)</td>
<td>72.1 (9.87)</td>
<td>72.3 (9.07)</td>
</tr>
<tr>
<td>Gender (% Male)</td>
<td>9 (64%)</td>
<td>8 (50%)</td>
<td>4 (40%)</td>
</tr>
<tr>
<td>BMI</td>
<td>27.7 (5.53)</td>
<td>31.9 (7.95)</td>
<td>32.5 (7.04)</td>
</tr>
<tr>
<td>FVC</td>
<td>3.30 (1.03)</td>
<td>2.92 (1.29)</td>
<td>3.84 (0.84)</td>
</tr>
<tr>
<td>FEV1</td>
<td>1.88 (0.81)</td>
<td>1.51 (0.67)</td>
<td>2.12 (0.79)</td>
</tr>
</tbody>
</table>

Table 1: Patient characteristics at enrolment.

Patient characteristics within disease groups separated by NTM infection at recruitment. All characteristics were taken at recruitment for the NTM group with multiple visits. Values in brackets show SD in groups unless stated otherwise.

Protein profiles differ between diseases.

The sputum proteome was compared between groups through PCA. Initially visit 1 samples were analysed meaning that we included only one sample per patient, and all samples were prior to any NTM treatment. Figure 1 indicates diverse sputum proteomic profiles among patients with COPD, bronchiectasis and CF with a partially separation observed between COPD and CF or bronchiectasis. T-test with Benjamini & Hochberg correction for multiple comparisons show no significant changes in protein abundance between patients with CF or bronchiectasis. KLK1 was upregulated in bronchiectasis compared to COPD. A total of 12 proteins were upregulated including MPO, AZU1, CTSG, CAT and RNASE3 in CF compared to COPD, with 21 proteins downregulated including SCGB1A1, IGFBP2, SFTPB, GC and CFD (full list of significantly changed proteins in e-Table 1). Linear regression analysis confirmed that the effects of disease on major protein differences between CF and COPD were independent of the effects of *P. aeruginosa* (e-Table 2). These demonstrate distinct proteomic profiles between disease groups.

*Pseudomonas aeruginosa* drives protein profile
The most frequently isolated typical bacterial pathogens in this population were *Haemophilus influenzae*, *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Streptococcus pneumoniae*. For the purposes of analysis *P. aeruginosa* was compared to other gram-negative (predominantly *H. influenzae*) and gram-positive (predominantly *S. pneumoniae* and *S. aureus*) bacteria (Figure 2). We identified significant differences in protein profiles between the *P. aeruginosa* infected group and the un-infected group (Figure 2C), but not between non-*Pseudomonas* gram-negative infections. In the gram-positive infections LGALS3BP, SCGB1A1, APOH and FOLR1 were upregulated compared to *P. aeruginosa* infections. 47 proteins were significantly up-regulated in the presence of a *P. aeruginosa* infection compared to no infection (full list of significantly changed proteins in e-Table 3). They are largely neutrophil derived including neutrophil elastase, azurocidin, MPO and S100A9/A8. Patients with *P. aeruginosa* infections were split between the CF (6) and bronchiectasis (4) groups with no significant differences between them, therefore indicating that underlying disease is unlikely to be a strong influence in the differences observed.

**Protein profiles of NTM infections**

Following characterising differences between disease groups we looked to determine differences in proteomic profile associated with NTM infections (Figure 3). Initially, we compared patients with or without NTM infections in principal component 1 and 2 from figure 1. Here we saw no clear distinction between patients with or without NTM infections independent of disease group. To investigate the impact of treatment and NTM status on sputum proteomic profiles we compared protein abundance between patients’ first visit, where they had active NTM infection, and sputum obtained following sputum culture conversion for the 11 patients with repeat samples. One patient did not achieve sputum culture conversion and therefore had positive samples at all 5 visits, so their final positive sample was included in analysis. Principal component analysis of these patients’ proteomic profiles indicate no clear association between NTM status and sputum proteome. This is supported by no significant difference in individual protein abundance between NTM and non-NTM groups (data not shown).

**Severity of NTM disease**

A comparison of the proteome between severe (n=6) and non-severe (n=9) NTM cases prior to treatment demonstrated HPX, TF and A1BG were significantly downregulated in patients with severe disease (Figure 4). HPX and TF are iron chelating proteins. The most up and down-regulated proteins are shown in e-Table 4.
NTM treatment response

As NTM infections cause a heterogeneous disease, with variations between patients’ lung damage and clinical improvement over time, we looked at how the proteome of individual patients changed over the 5 visits in relation to their clinical improvement. Regimens used were rifampicin, ethambutol and macrolide (6 patients), IV therapy followed by oral minocycline, ciprofloxacin, clarithromycin and nebulized amikacin (3 patients), IV therapy followed by oral minocycline, clarithromycin, clofazamine and nebulized amikacin (1 patient) and IV therapy followed by oral doxycycline, clofazimine, azithromycin and nebulized meropenem (1 patient). Regimens could be modified as clinically indicated. Based on the association between HPX, TF and A1BG and disease severity we initially hypothesised that these markers would decrease with treatment. Trajectories of these markers are shown in figure 4 and show trends towards a decrease in these marker abundances over time in less severe disease, but an increase in severe disease. A1BG shows a significant association of protein abundance over time (ANOVA of linear mixed effects model, visit number, F value=0.0445) although no significant differences are seen for TF and HPX.

We then analysed patient treatment over time in relation to multiple proteins using principal component 1, determined through inclusion of all NTM patients proteomes over repeat visits. This component’s positive values are largely driven by markers of the adaptive immune response (IGHG1/IGHM), acute phase response (ORM1, A2M, FGB, C3) and neutrophil proteins including ELANE, MPO, AZU1 and negative values are driven by proteins including PSMA4 and NPEPPS-NPEPPSL1 (e-Table 5) associated with proteasome degradation. The neutrophil proteins in particular have been related to symptoms and are important prognostic markers in COPD, bronchiectasis and CF. Figure 5 shows the sputum protein profiles over repeat visits during treatment. Three patients showed a clear improvement in inflammatory status reflected in a decreasing abundance of pro-inflammatory proteins in at least 4 of 5 on treatment samples. All three patients also reported a clear improvement in symptoms during this period. In contrast, 3 patients displayed deteriorating sputum proteomic profiles over at least 4 of 5 visits while on treatment, with increasing pro-inflammatory protein levels. In 2 patients this correlated with no perceived clinical improvement. Participant 38, also showed no improvement in proteomic profile and this patient remains culture positive for NTM throughout the study despite treatment. The other participants showed inconsistent protein profiles with no trend over time.

Discussion
Sputum proteomics is a powerful tool for biomarker discovery in airways disease. Proteomic studies previously have identified calprotectin as a biomarker of disease severity and exacerbation in CF, neutrophil extracellular traps and SPLUNC1 in bronchiectasis and multiple candidate biomarkers in chronic obstructive pulmonary disease.26,20–23 By profiling more than 600 sputum proteins we aimed to identify candidate biomarkers of NTM lung disease. We identify that protein profiles from sputum are heavily influenced by the underlying disease and the presence of bacterial co-infections and particularly P. aeruginosa. This identifies a critical challenge for any future work attempting to develop biomarkers for NTM. NTM may arise in the context of multiple chronic lung conditions including CF, bronchiectasis, COPD, asthma and immunosuppressive conditions.1 Since each of these conditions, and the co-infection with gram-negative and gram-positive bacteria that accompany them, are associated with highly variable levels of airway inflammation,24,25 microbial dysbiosis,26,27 and immune/epithelial dysfunction28, taking into account the background condition and associated infections appears to be crucial. Despite these complexities we identify 3 candidate biomarkers of NTM disease severity as well as demonstrating relationships with NTM treatment response.

2 proteins (TF and HPX) downregulated in non-severe disease are associated with iron chelation29, with biological plausibility as iron is an essential co-factor for NTM metabolism and growth and both have previously been identified in tuberculous granulomas.30,31 As this was not a focus of our study we do now know if these have had an influence on free iron levels, but it is plausible that reduced iron chelation would increase free iron and promote NTM growth and virulence. Additionally, these proteins show promising trends in abundance associated with disease severity and treatment response, although this study was only sufficiently powered to show significant treatment response for A1BG. These biomarkers now require independent validation in larger cohorts.

We recently showed a strong association between neutrophil extracellular trap (NET) derived protein signatures and clinical outcomes in bronchiectasis.16 A high degree of overlap was observed in inflammatory profiles between CF, bronchiectasis and COPD and indeed no significant differences between CF and bronchiectasis were identified. While disease labels give an indication of how patients have developed lung damage our data supports the emerging paradigm that patterns of airway inflammation are shared between different diseases, and that treatments for neutrophilic disease such as macrolide antibiotics and emerging anti-inflammatories, or eosinophilic disease (inhaled corticosteroids, anti-IL5/IL5R monoclonal antibodies) may be targeted through a so called “treatable traits” approach that is agnostic to disease label.34–36
Biomarkers are urgently needed for NTM lung disease due to the complexity and heterogeneity of the condition. Diagnostic and treatment decisions including identifying when NTM is present, deciding when to initiate treatment, how to define treatment response, duration of treatment and how to prevent recurrence are among multiple areas of uncertainty that could be supported by biomarkers. There have been previous studies designed to identify potential biomarkers of NTM lung disease. Cowman et al, studied 25 patients with NTM lung disease and 27 controls without NTM and identifies a total of 213 transcripts that were significantly differentially expressed by gene expression analysis in NTM lung disease. Prominent among these were interferon-gamma and Th1 responses which were downregulated in NTM lung disease. Whether these represent biomarkers of NTM-infection or innate susceptibility to disease is not known. Smaller studies have found differences in peripheral blood micro-RNA between NTM lung disease and controls. A study of 187 patients with M. avium lung disease found serum levels of Krebs von den Lungen-6 (KL-6) were associated with a diagnosis of NTM-and higher levels were associated with disease progression. Studies in peripheral blood may, however, be remote from the primary site of infection which is the airway.

A limitation of the majority of NTM biomarker studies to date, including the present study, is sample size. NTM is a rare disease that presents insidiously, restricting the sample sizes available for study. In small studies, a large “signal” is required to overcome biological and experimental “noise”. The high sensitivity and accuracy of sputum proteomics means it is possible to detect very large differences between groups based on small samples sizes – for example, in the study by Keir et al, 39 significantly differentially expressed proteins were identified with a sample size of only 20 patients. We therefore anticipated that we could observe proteomic changes with the prolonged guideline recommended antibiotic therapy for NTM. We, however, identified no statistically significant differences in protein profiles with treatment, but trajectories that suggested some patients who achieved a reduction in acute phase, adaptive immune and neutrophilic inflammatory pathways also experienced a clinical improvement while those that had no change or a deterioration in protein profile did not experience clinical improvement. The assessment of clinical response was subjective which is a limitation although assessment was blinded to protein profiles. It has proven challenging to demonstrate symptomatic benefits with antibiotic treatment for NTM since NTM therapies often have associated adverse effects. As an example of this issue, inhaled liposomal amikacin significant improved sputum culture conversion rates compared to guideline based treatment alone in the CONVERT trial, but symptoms measured using the SGRQ score were not improved and in fact symptoms were numerically better in the guideline based treatment group (mean difference 3.8
In our prior studies symptoms were strongly linked to bacterial load of conventional bacteria and neutrophilic inflammation, and therefore improved by treatment to reduce bacterial loads. It is possible, therefore that under certain circumstances treating *P. aeruginosa* or other pathogens is more likely to reduce lung inflammation and generate a symptomatic improvement in patients than treating NTM. Conventional culture was not performed routinely during follow-up in our study and it is possible that the observed clinical responses were due to the effect of antibiotics on other colonizing pathogens rather than specific effects on NTM. A limitation of this study is the use of sputum which while an accessible matrix used for diagnosis and monitoring of NTM patients may not reflect inflammatory processes occurring within inflammatory cells and the pulmonary parenchyma which are important sites of NTM infection. A strength of our study is the multicentre design but much larger samples sizes will be required in future with biological samples obtained using standardized methodology to realistically develop a robust biomarker of set of biomarkers for NTM-lung disease. A multicentre, multinational registry linked to a prospectively collected biosamples similar to the EMBARC-BRIDGE biobank (NCT03791086) for bronchiectasis and focused on NTM would achieve this objective.

**Interpretation**

In summary, sputum proteomics demonstrates differences in inflammatory profiles between individuals with CF, bronchiectasis and COPD and is strongly influenced by the presence of *P. aeruginosa* and other typical bacterial infections. Consequently, specific sputum proteins biomarkers could not be identified independently from the underlying chest disease and co-infections. Future studies seeking NTM biomarkers will require to account for these factors and include larger sample sizes.

**Acknowledgments**

RCH had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. RCH, JTJH and JDC carried out data analysis and interpretation. The manuscript was written by JDC and RCH. RCH, JTJH, AKB, HRK, HE, WC, MM, ML and JDC contributed substantially to the study design, data generation and analysis and revision of the manuscript.

**Financial/nonfinancial disclosure:** 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 GSK/British Lung Foundation Chair of Respiratory Research and a Scottish Senior Fellowship from the Chief Scientist Office. RCH is supported by a scholarship from the Medical Research Foundation National PhD Training Programme in Antimicrobial Resistance Research (MRF-145-0004-TPG-AVISO) and the AMR cross-council funding from the MRC to the SHIELD consortium “Optimising Innate Host Defence to Combat Antimicrobial Resistance” MRNO2995X/1.
Role of sponsors: The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

Take home point:

Study Question: Can proteomics identify biomarkers of infection, disease severity or treatment response in non-tuberculous Mycobacterial lung disease?

Results: We show that underlying disease and Pseudomonas aeruginosa infection are key drivers of sputum protein profile, but analysis of disease severity identifies reduced iron chelation markers in severe NTM disease and reductions in immune response proteins were associated with response to NTM therapy.

Interpretation: Sputum proteomics is a promising technique to identify NTM biomarkers but future studies will require to take underlying disease and “typical” bacterial infections into account.

Figure Legends:

Figure 1: Influence of underlying disease on sputum proteome
Sputum proteomic analysis of 693 proteins from visit 1 for all participants in NTM and non-NTM disease groups. A) Principal component analysis with groups separated by disease, red square= BE, blue triangle= CF, green x =COPD. Ellipses represent core area for each group with 68% CI. B) Top 15 contributing variables to principal components 1 and 2 used in plot A. C) Volcano plots of mean change in protein expression between CF vs. COPD. p-values show T-test between 2 disease groups with Benjamini & Hochberg correction. 21= EEF1A1P5 EEF1A1 EEF1A2, 605= PRDX1. Significantly upregulated proteins (p<0.05) are represented in red. Full list of upregulated proteins in e-Table 1.

Figure 2: Secondary infections influence proteomic profile of patients with NTM infections
Sputum proteomic analysis of 693 proteins from visit 1 for all participants in NTM and non-NTM disease groups. A) Principal component analysis with groups separated by presence of non-NTM co-infections, red square= no co-infection, blue triangle= Pseudomonas spp, green x =gram-negative, purple diamond=gram-positive, orange circle=mixed co-infection. Ellipses represent core area for each group with 68% CI. B) Top 15 contributing variables to principal components 1 and 2 used in plot A (same as figure 1B). C) Volcano plots of mean change in protein expression between no co-infection vs. Pseudomonas co-infection. p-values show T-test between 2 infection groups with Benjamini & Hochberg correction. 417= HIST1H1E HIST1H1C HIST1H1D HIST1H1A. Significantly upregulated proteins (p<0.05) are represented in red. Full list of upregulated proteins in e-Table 3.

Figure 3: Sputum proteomic profiles are unaffected by NTM status and treatment
PCA of 693 proteins from visits in control and NTM infection groups. Groups are divided by NTM status with red circles= control, blue cross= NTM group. Ellipses show grouping with 68% CI. B) Top 15 loading variables of PC1 and PC2 also shown in Figure 1A. C) PCA of NTM group with repeat visits sputum proteins. 2 sputum proteomes were selected for each patient based on NTM culture. No differences were observed between samples that are culture positive and the same patients samples after culture conversion. If no samples were negative their final sample is included. Grouping by NTM culture, red=negative, blue=positive. Numbers indicate the patient ID with dashed lines representing paired data. D) Top 15 loading variables of PC1 and PC2 in C.
**Figure 4: The influence of disease severity on sputum proteomic profiles**

A) Sputum proteomic analysis of 693 proteins including repeat visits from all participants in NTM disease groups. Volcano plots of mean change in protein expression between severe disease and non-severe disease groups. p-values show T-test between 2 infection groups with Benjamini & Hochberg correction. Significantly upregulated proteins (p<0.05) are represented in red. Proteins with p-values <1 are shown in e-Table 4. B) Relative abundance of significantly upregulated proteins from A over the 5 patient visits from sputum samples of patients with NTM infections. Red= A1BG, blue= HPX, green= TF, solid line, square=non-severe disease, dashed line, cross=severe disease. Trend lines show linear models fitted to these data.

**Figure 5: Patients’ sputum proteomic profile over time in relation to clinical improvement**

Principal component 1 from PCA of 693 proteins from sputum samples of 11 patients with NTM infections with repeat visits. Numbers indicate patient ID-sample number. 5 samples (visit 1- visit 5) from each patient are separated over time. Clinical improvement is defined by the managing clinician blinded to protein profile results, red= no improvement, blue=improvement. Yellow dashed line= 2 SD of PC1. Large arrows indicate directional trends in the patient samples between the highest and lowest sample, blue=improvement, red=decline. Black box highlights one patient who was refractory to treatment and did not achieve culture conversion over the 5 visits. Contributing protein coordinates are in eTable 5.
References
