Contents lists available at ScienceDirect

# Respiratory Medicine

journal homepage: www.elsevier.com/locate/rmed

Original Research

# The relationship between lung disease severity and the sputum proteome in cystic fibrosis

Rosemary E. Maher [a], Emma Barrett [b,c], Robert J. Beynon [a], Victoria M. Harman [a], Andrew M. Jones [d,g], Paul S. McNamara [e], Jaclyn A. Smith [f,g], Robert W. Lord [d,g,*]

[a] Centre for Proteome Research, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, United Kingdom
[b] Department of Research and Innovation, Manchester University NHS Foundation Trust, Manchester, United Kingdom
[c] Centre for Biostatistics, Faculty of Biology Medicine and Health, University of Manchester, Manchester, United Kingdom
[d] Manchester Adult CF Centre, Manchester University NHS Foundation Trust, Southmoor Road, Wythenshawe, Manchester, M23 9LT, United Kingdom
[e] Department of Child Health (University of Liverpool), Institute in the Park, Alder Hey Children's Hospital, Eaton Rd, Liverpool, L12 2AP, United Kingdom
[f] Department of Respiratory Medicine, Manchester University NHS Foundation Trust, Southmoor Road, Wythenshawe, Manchester, M23 9LT, United Kingdom
[g] Division of Infection, Immunity and Respiratory Medicine, University of Manchester, Manchester, M13 9PL, United Kingdom

A B S T R A C T

*Background:* Proteomics can reveal molecular pathways of disease and provide translational perspectives to inform clinical decision making. Although several studies have previously reported the cystic fibrosis airway proteome, the relationship with severity of lung disease has not been characterised. The objectives of this observational study were to investigate differences in the CF sputum proteome associated with disease severity and identify potential markers of disease with translational potential.
*Methods:* Sputum samples from healthy volunteers and cystic fibrosis subjects (some prescribed modulator therapies) were analysed using liquid-chromatography tandem mass spectrometry. Severity of lung disease was based on baseline spirometry (percentage predicted forced expiratory volume in 1 s, FEV1%).
*Results:* Multiple sputum proteins (108 increased; 202 decreased) were differentially expressed in CF (n = 38) and healthy volunteers (n = 32). Using principal component analysis and hierarchical clustering, differences in sputum proteome were observed associated with progressive lung function impairment. In CF subjects, baseline FEV1% correlated with 87 proteins (positive correlation n = 20, negative n = 67); most were either neutrophil derived, or opposed neutrophil-driven oxidant and protease activity.
*Conclusion:* Predictable and quantifiable changes in the CF sputum proteome occurred associated with progressive lung function impairment, some of which might have value as markers of disease severity in CF sputum. Further work validating these markers in other patient cohorts and exploring their clinical utility is needed.

## 1. Introduction

The destructive bronchiectasis seen in cystic fibrosis (CF) lung disease leads to progressive respiratory failure and premature death [1]. Dysfunction of the cystic fibrosis transmembrane conductance regulator (CFTR) protein results in impaired mucociliary clearance, that coupled with disturbed immune and epithelial cell behaviour predisposes to chronic bacterial infection and persistent inflammation in the CF airway [2]. An important advance over the past decade has been the development of CFTR modulators that address the underlying defect in CFTR function, but not everyone with CF can take these new medications and

they are not universally available. Improved understanding of the processes underlying CF lung disease could aid the discovery of new targets for both novel therapeutics and disease monitoring.

Proteomics, the study of large-scale changes in protein expression within biological systems, can reveal molecular pathways of disease and provide translational perspectives to inform clinical decision making [3]. In CF, proteomics has defined protein abundance changes in epithelial cells cultured in vitro or from animal models [4,5]. Proteomic analysis of CF sputum and bronchoalveolar lavage fluid samples has established that the CF respiratory tract proteome is distinct from healthy controls, with the differences largely driven by upregulation of

neutrophil-related immune and inflammatory responses, as well as differences in a number of pathways including those related to protease activity, oxidative stress and actin cytoskeleton rearrangement [6–9]. Individual proteins relate to severity of lung disease as determined by baseline percentage predicted forced expiratory volume in 1 s (FEV1%) [7,10–17]. However, the relationship between severity of lung disease and the global proteome is less clear.

We hypothesised that there will be differences in sputum proteome that associate with severity of CF lung disease. Sputum samples from adults and healthy controls were compared using quantitative comparative proteomics. The proteome differences that aligned with increased severity of disease were defined.

## 2. Materials and methods

### 2.1. Subjects and study design

This cross-sectional observational single-centre study was delivered at a large UK specialist CF centre and conducted in line with ethical approval granted by The Greater Manchester West Research Ethics Committee (15/NW/0655). All subjects provided written informed consent.

CF subjects were aged over 18 years, had a confirmed diagnosis using genetic testing and/or sweat testing with typical phenotypic features, and were deemed clinically stable by the medical team when samples were collected. Exclusion criteria were pregnancy and lung transplantation. Baseline lung function was obtained on the same day samples were collected. Disease severity in CF subjects was based on baseline FEV1, such that mild disease was defined as FEV1>70%; moderate 40-70%; and severe <40%. The control group comprised adult non-smokers with no significant respiratory or gastro-intestinal past medical history and no evidence of airway obstruction (FEV1% predicted ≥80% and FEV1/FVC ratio ≥0.7).

### 2.2. Sputum collection and processing

Spontaneous sputum samples were collected and stored from CF subjects. Induced sputum samples were collected from healthy control subjects using 7% hypertonic saline [18]. The sputum plugs were manually separated from the salivary component of samples and then stored at −30 °C prior to analysis. They were then thawed and dithiothreitol added. Following three periods of agitation on a vortex mixer and straining to remove large particulates, samples were centrifuged. The recovered supernatant fraction was aliquoted and stored at −80 °C.

### 2.3. Proteomic analysis

Details of the proteomic analysis are provided in the supplementary material. After thawing, supernatant fractions were treated with RapiGestTM SF Surfactant (Waters, Wilmslow, UK), reduced with dithiothreitol, and alkylated with iodoacetamide. Proteins were digested overnight with trypsin at 37 °C. Digestion was stopped by addition of trifluoroacetic acid. The samples were centrifuged to remove particulates. Tryptic peptides within samples were analysed by liquid chromatography mass spectrometry using an Ultimate 3000 nano system coupled to QExactive-HF mass (Thermo Fisher Scientific, Hemel Hempstead, UK).

### 2.4. Protein identification

Run alignment and peak picking was carried out using Progenesis QI for proteomics v4 (Waters, Wilmslow, UK), while database searching used the MASCOT search engine version 2.6 (Matrix Science, London, UK). Samples were searched against a database comprising all human sequences in the UniProt UniHumanReviewed database (updated 12/08/2020; 20,356 sequences; 11,357,197 residues) using trypsin as the specified enzyme, carbamidomethylation of cysteine as fixed modification, methionine oxidation as variable modification and one trypsin missed cleavage, precursor and fragment ion error tolerances were set to 10 ppm and 0.01 Da, respectively. The false discovery rate (FDR) was calculated using the decoy database tool in MASCOT. Only proteins identified with an FDR <1% were accepted.

### 2.5. Statistical analysis

All statistical analyses were undertaken within the R environment (v.1.2.5033). Mann-Whitney U tests compared protein abundance of in CF and healthy control subjects, with correction for multiple comparisons [19]. Proteomic data were further analysed by hierarchical clustering and principal component analysis (PCA), which are two commonly used unsupervised statistical techniques for investigation of study populations with biological datasets. Hierarchical clustering analysis attempts to group subjects based on similarity of their datasets into clusters, which are then displayed using dendrograms and heat maps [20]. Principal component analysis is a different approach that allows interpretation of the dataset by reducing the number of variables whilst retaining as much of the variance as possible. To do so existing variables undergo linear transform to form new variables called principal components. The first principle component captures the largest amount of variance, with each subsequent principal component responsible for progressively less variance [21]. The relationship between the first two principal components and FEV1% was examined using Spearman's rank analysis.

Correlations between protein abundance and FEV1% were examined using Spearman's rank analysis. Gene set enrichment analysis was performed on the proteins relating to lung disease severity, as determined by those correlating with FEV1% (p < 0.05). This approach was originally designed for genomic experiments but has been adapted to identify overrepresented protein sets. It involves calculating an enrichment score which then determines a statistical significance, which is then adjusted for multiple hypothesis testing [22]. Using the ShinyGO server important molecular pathways and biological functions were identified that related to lung disease severity [23].

## 3. Results

### 3.1. Clinical and demographic characteristics

Table 1 shows the clinical and demographic characteristics for healthy control and CF subjects, the latter grouped based on severity of lung disease. Baseline characteristics were generally well-balanced between CF and healthy controls with some differences in sex between groups, although these do not appear significant (Figure E1 in online data supplement). Individual characteristics are shown in the supplementary material (Table E1 in online data supplement).

### 3.2. Differences in the proteome between CF and healthy controls

The total protein concentration of sputum was increased in CF (mean ± SD, 4.78 ± 2.56 mg/mL) compared to healthy controls (1.72 ± 0.98 mg/mL). Protein identification was based on at least two unique peptides. This resulted in 440 proteins being detected and of these, 415 were common to all 70 subjects. Of these 415 proteins, 310 differed in abundance between CF patients and healthy controls (adjusted p-value ≤ 0.05); 108 proteins were increased in CF samples and 202 were decreased (Fig. 1).

Hierarchical clustering analysis was undertaken using the 20 proteins that discriminated most between CF and healthy control samples based on the lowest adjusted p-values (Fig. 2). Generally, proteins more abundant in CF samples were neutrophil derived, such as elastase and myeloperoxidase, whereas those less abundant had immunomodulatory properties, antiprotease or antioxidant activity, or were acute phase

**Table 1**
Clinical and demographic characteristics of healthy control and CF subjects.

| | Healthy control (n = 32) | Mild CF (n = 4) | Moderate CF (n = 25) | Severe CF (n = 9) |
|---|---|---|---|---|
| **Age, mean (SD), yr** | 36 (10) | 30 (8) | 31 (8) | 32 (9) |
| **Baseline FEV1%, mean (SD)** | 99 (36) | 87 (7) | 52 (9) | 30 (7) |
| **Body mass index, mean (SD), kg/m²** | 24.7 (4.6) | 24.3 (5.2) | 22.3 (2.5) | 21.0 (3.3) |
| **Sex, Male, n (%)** | 8 (25) | 4 (100) | 18 (68) | 5 (56) |
| **Homozygote Phe508del, n (%)** | NA | 2 (50) | 15 (60) | 7 (78) |
| **CF related diabetes n (%)** | NA | 1 (25) | 14 (56) | 4 (44) |
| **CFTR modulation** | | | | |
| Any, n (%) | NA | 1(25) | 2 (8) | 3 (33) |
| Ivacaftor, n (%) | NA | 1 (25) | 1 (4) | 1 (11) |
| Ivacaftor/ lumacaftor, n (%) | NA | 0 (0) | 1 (4) | 2 (22) |
| **Prednisolone** | NA | 1 (25) | 4 (16) | 1 (11) |
| **Long-term macrolide n (%)** | NA | 4 (100) | 22 (88) | 9 (100) |
| **Chronic *Pseudomonas aeruginosa*, n (%)** | NA | 1 (25) | 19 (76) | 7 (78) |
| **Chronic *Burkholderia cepacia* complex n (%)** | NA | 1 (25) | 2 (8) | 3 (33) |
| **Chronic Non-tuberculous mycobacterium n(%)** | NA | 0 (0) | 1 (4) | 0 (0) |

proteins. There was clear resolution of CF and healthy controls with the exception of one CF subject (CF021), who had mild lung disease (baseline FEV1% 96%) and clustered with the healthy control subjects.

One healthy control (HV903) sample was flagged as an extreme outlier by PCA and upon inspection derived from a low abundant LC-MS/MS trace, requiring a normalisation gain factor of 12.3 (median normalisation gain factor 1.1) and was therefore removed from the dataset. Separation of CF samples from healthy controls was evident by principal component analysis and largely driven by differences in the first principal component (PC1, Fig. 3a), which explained 22% of the total variance. Unsurprisingly the 20 proteins making the largest contribution to PC1 included 13 proteins previously flagged as having the most statistically significant expression profile (see Fig. 4). The seven additional proteins related to immune function and inflammation. Neither the first nor the second principal component appear to relate to sex of subject (see Figure E1 in the online data supplement).

*3.3. Relationship between sputum proteome and severity of lung disease*

Both hierarchical clustering and PCA demonstrated a continuum from healthy control, through mild, to moderate and severe lung disease based on baseline FEV1 (Figs. 2 and 3b). Most CF subjects with mild lung disease had proteomes closely resembling healthy controls (the one CF subject who co-localised with those with severe disease suffered a respiratory exacerbation within a month of the sputum sample being taken). CF subjects with mild disease overlapped with those with moderate disease who in turn overlapped with those with severe disease (Figs. 2 and 3b). Analysing FEV1% as a continuous variable revealed a moderate correlation with PC1 (Spearman's rank, P = 0.005, R = 0.444) (Fig. 3c), but not PC2 (P = 0.291, R = 0.076) (Fig. 3d). The nature of the proteins contributing to PC1 from the PCA and hierarchical clustering suggest that predictable and quantifiable increases in immune function and inflammation occur as lung function worsens. The shifts in proteome are again heavily influenced by neutrophil-derived proteins.
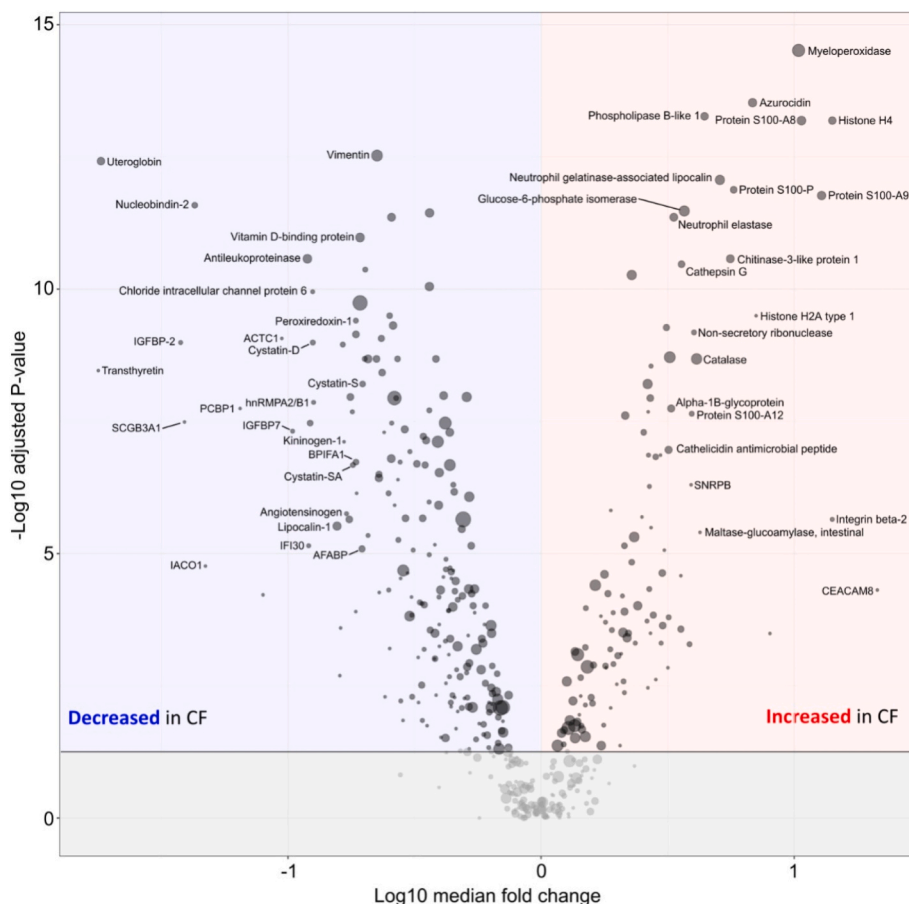


**Fig. 1. Quantitative profiling of cystic fibrosis and healthy volunteer sputum proteomes.** A total of 69 samples (38 CF and 31 healthy control) were analysed by proteomics. Protein abundances, assessed by label-free quantitative proteomics are displayed as a volcano plot. Symbols corresponding to proteins with the greatest statistical confidence and highest fold change are labelled. The symbol size maps to the number of peptides used for quantification with the smallest being two unique peptides.
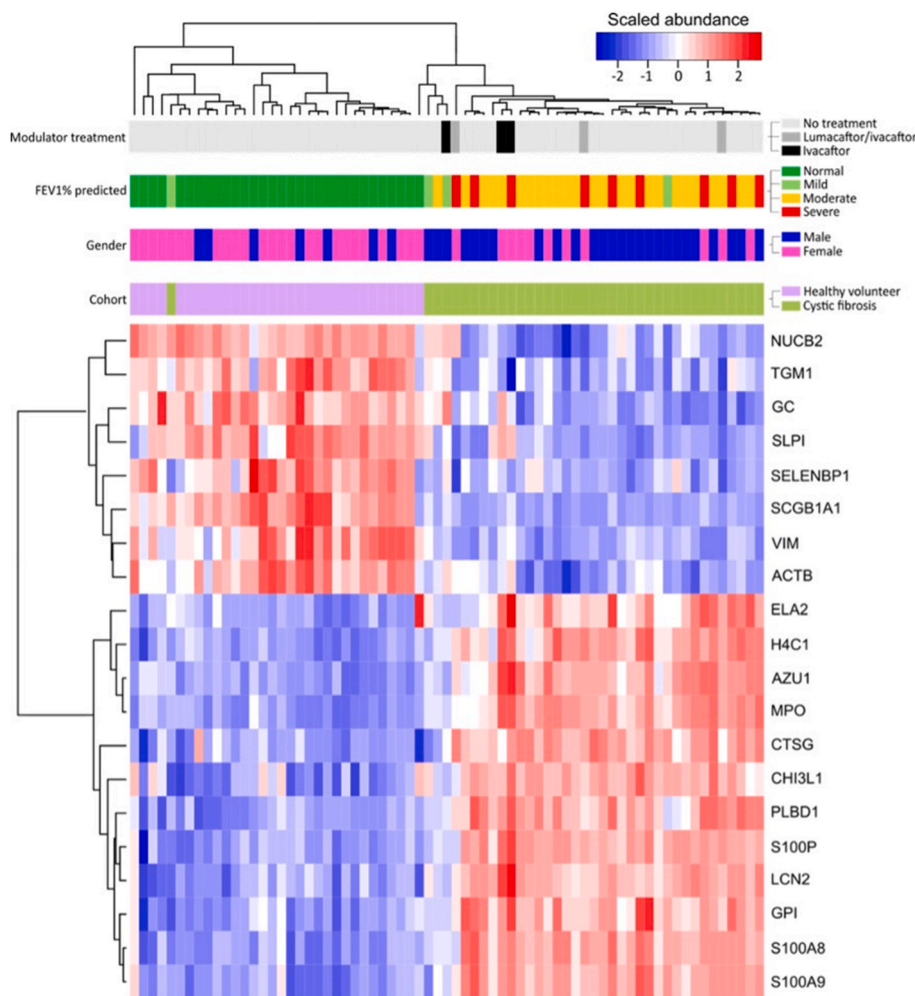
**Fig. 2. Hierarchical clustering using the proteins with the greatest discrimination between CF and healthy control samples.** The 20 proteins that discriminated most between CF and healthy control samples were determined based on the lowest adjusted p-values. Specific characteristics of the subjects are mapped at the head of the heatmap. *Definition of abbreviations*: NUCB2 = nucleobindin-2 precursor; TGM1 = transglutaminase 1; GC = gc globulin (vitamin D-binding protein): SLPI = secretory leukocyte peptidase inhibitor; SELENBP1 = selenium-binding protein 1 (methanethiol oxidase); SCGB1A1 = secretoglobin family 1A member 1 (uteroglobin); VIM = vimentin; ACTB = actin beta (beta actin); ELA2 = elastase 2 (neutrophil elastase); H4C1 = H4 clustered histone 1 (histone H4); AZU1 = azurocidin 1; MPO = myeloperoxidase; CTSG = cathepsin G; CHI3L1 = chitinase-3-like protein 1; PLBD1 = Phospholipase B domain containing 1 (phospholipase B-like 1); S100P = S100 calcium-binding protein P; LCN2 = ; GPI = ; S100A8 = S100 calcium-binding protein A8; S100A9 = S100 calcium-binding protein A9.

Of the 87 proteins that correlated with baseline FEV1% (p < 0.05), 67 increased with worsening lung function, and a further 20 showed the inverse relationship (see Table E2 in the online data supplement). Many of the proteins that increased with severity were also those that showed the greatest discrimination between CF and healthy control samples (10 from the top 20 in Fig. 2) and contributed strongly to PC1 in the PCA (nine from the top 20 in Figure E1 in the online data supplement). Of the proteins that increased with severity (Fig. 6) neutrophil elastase, histones H2A and H2B, catalase and glucose-6-phosphate isomerase are components of neutrophil extracellular traps (NETs) [24]. Triose phosphate isomerase, peptidoglycan recognition protein 1 and phospholipase-B-like-1 are also neutrophil-derived and implicated in the innate immune response [25–27]. Levels of alpha-amylase 1 were lower in patients with severe lung disease, which likely reflects their ability to expectorate sputum samples with minimal salivary contamination. The molecular functions and biological processes associated with all 87 proteins that correlated with baseline lung function were mapped using Gene Ontology Resource, and were in keeping with progressively increasing neutrophil activity and greater disturbance of imbalances in protease/anti-protease and oxidant/antioxidant states (see Fig. 5).

### 3.4. CFTR modulation and the sputum proteome

Hierarchical clustering analysis using all 440 proteins revealed that five CF subjects co-locate with the healthy controls (see Figure E2 in the online data supplement). Two had mild disease (CF014, CF021), and three were prescribed Ivacaftor (Kalydeco) (mild disease [CF032], moderate [CF045] and severe [CF055]). The genotypes of those

prescribed ivacaftor were heterozygote for Gly551asp (CF032 and CF045) and Arg117His (CF055). All CF subjects taking ivacaftor co-located with healthy controls irrespective of underlying lung function impairment. PCA of CF subjects prescribed ivacaftor positioned away from those with similar lung disease severity and closer to healthy controls (Fig. 3). These findings were not replicated for CF subjects taking ivacaftor/lumacaftor (Orkambi) (CF029, CF052 and CF056).

## 4. Discussion

We describe, for the first time, differences across the sputum proteome that become more pronounced with worsening CF lung disease. These large-scale differences in protein expression relate to increased neutrophil activity and imbalances in protease/anti-protease and oxidant/antioxidant states, leading to a grossly abnormal environment within CF airways with significant potential for injury. These findings build upon previous work using mass spectrometry in the analysis of sputum and bronchoalveolar lavage from CF subjects [7–9]. In using what is a uniquely large dataset for sputum proteomic studies in CF, we have demonstrated utility of this approach as a technique capable of providing insights into disease mechanisms, and identifying potential biomarkers capable of classifying disease severity and activity.

In CF, microbial dysbiosis and hyperinflammatory responses within immune and epithelial cells cause an influx of neutrophils in the airway. In accord with previous studies, we have shown differences in the sputum proteome between CF and control samples consistent with heightened neutrophil activity (Fig. 1) [6–8]. Of note, we also show that the sputum proteome becomes increasingly laden with
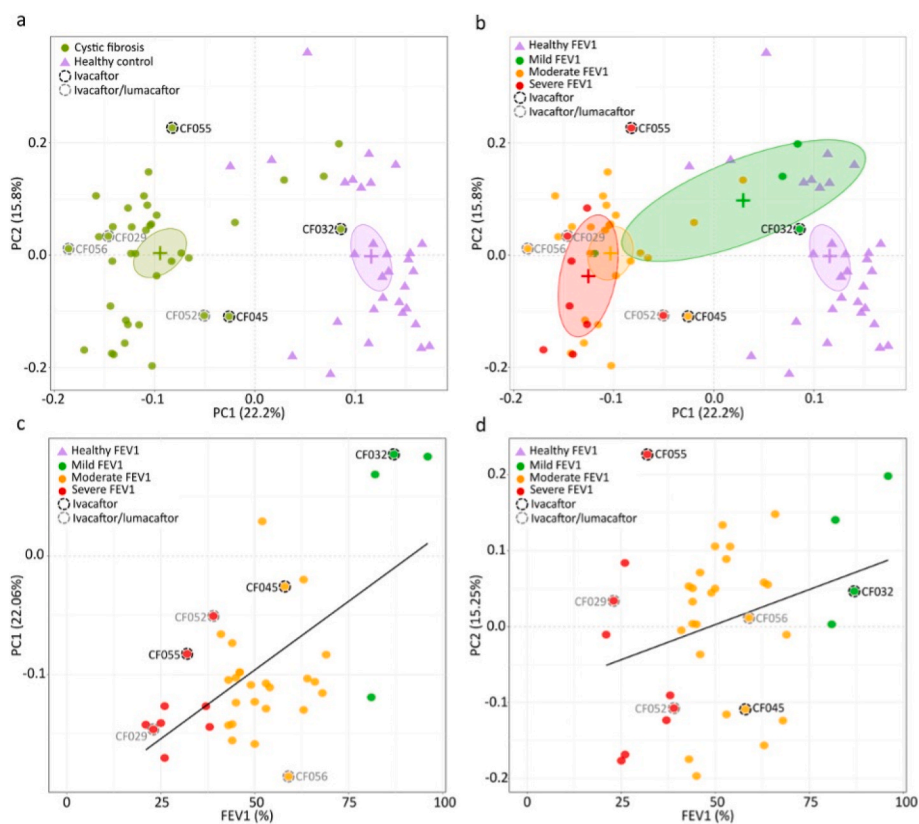
**Fig. 3. Principal component analysis of sputum protein profiles.** (a) Score plot with CF subjects and healthy controls subjects labelled. Each data point is an individual subject with 95% confidence intervals overlayed (superimposed ellipses, centroid marked by cross). The subjects on CFTR modulation are labelled. (b) Score plot with labelling of CF subjects based on severity of lung disease as defined by FEV1%: mild >70%; moderate 40-70%; severe <40%. (c) Correlation between principal component 1 and baseline lung function (d) Correlation between PC2 and baseline lung function.

neutrophil-derived proteins as lung function impairment progresses. Of these proteins, many are released in response to infection. Some have previously been shown to relate to CF lung disease severity or increase during pulmonary exacerbations, such as myeloperoxidase [13], neutrophil elastase [10,11], azurocidin [10], chitinase 3 like protein [15], protein S100-A8/A9 [16], cathepsin G [17] and matrix metallopeptidase 9 [13], whereas others are novel, such as, peptidoglycan recognition protein 1 and triosephosphate isomerase. Additional intracellular proteins known to be part of the core NET proteome [24], including several histones and their isomers, catalase and glucose 6 phosphate isomerase, were detected at high levels in CF samples. Some of these intracellular proteins become increasingly abundant with progressive lung function impairment likely indicating enhanced NET formation due to increased infective or inflammatory stimuli within the airway, and may also represent additional biomarkers in CF. High levels of NETs may damage CF lung tissue by increasing exposure to the proteases incorporated within their structure, through histones leading directly to lung injury, and/or through NETs themselves acting as pro-inflammatory stimuli [28]. These effects may worsen airway dysbiosis [29] and cause more neutrophilic inflammation [30].

The dysregulated inflammation within the CF airway is characterised by an abundance of neutrophil derived proteases and oxidants such as neutrophil elastase and myeloperoxidase, and a lack of anti-protease and antioxidant proteins, such as those belonging to the serpin [31], cystatin [32], and peroxiredoxin [33] protein families. Other proteins with immunomodulatory actions that may reduce inflammation within the lung, such as uteroglobin, vimentin, nucleobindin-2 and clusterin are also reduced [34–37]. Again, this disruption appears to become more pronounced with progressive lung function impairment. The relationship of these counterregulatory and immunomodulatory proteins to CF lung disease severity is less well described and for some, such as cystatin D and alpha-1 antichymotrypsin, this is the first time they have been reported. These protein differences may highlight pathways for future therapeutic intervention.

The novelty of this study is showing the relationship between lung disease severity and the global proteome i.e., the pattern of changes in the entire set of proteins, which has not been previously demonstrated. This goes beyond existing studies that have only correlated baseline lung function with selected proteins from global proteomic analyses [7,14]. That we can replicate existing findings, such as increased neutrophil activity and imbalances in protease/anti-protease and oxidant/antioxidant states validates our novel findings and supports the utility of a global proteomics in early translational work for understanding disease mechanisms and identifying biomarkers. This proteomic approach can also make dramatic savings in time and resource, as evidenced by our replication of a large body of work examining the relationship between individual proteins and CF lung disease that has spanned many experiments [10–17].

Alpha-amylase 1 levels were higher in patients with mild lung disease (Fig. 6). As alpha-amylase 1 is a highly abundant salivary protein its presence suggests a degree of salivary contamination [38]. This is interesting as CF patients with milder lung disease expectorate samples with less sputum and a larger salivary component. To ensure contamination was minimised there was a preparatory step whereby sputum plugs were visually separated from the salivary portion. We feel this makes a dilutional effect an unlikely explanation for why neutrophil-derived proteins were reduced in milder subjects. If that was the case, we would expect to see in milder lung disease other salivary proteins being more abundant, whereas in fact the differences are driven by reductions in what appear to be sputum proteins. However, the preparatory step maybe insufficient to remove all salivary residue which would explain the detection of alpha-amylase 1 in samples and possibly the association seen with lung disease severity. In future analyses an additional validation step, such as a differential cell count to calculate the percentage of squamous epithelial cells, could be used to further ensure that sputum samples were not excessively contaminated with saliva.

The predictable and quantifiable nature of the differences seen in the

**Fig. 4. Abundance profiles for the key altered proteins.** For the twenty proteins making the largest contribution to PC1, the abundances are plotted logarithmically, and stratified according to categorisation of severity. For each group, individual protein abundances, together with median and interquartile ranges are plotted. Definition of abbreviations: S100A9 = S100 calcium binding protein A9; S100A8 = 100 calcium binding protein A8; TTR = transthyretin; MPO = myeloperoxidase; CLU = clusterin; SCGB1A1 = secretoglobin family 1A member 1; SLPI = secretory leukocyte peptidase inhibitor; NUCB2 = nucleobindin-2; H4C1 = H4 clustered histone 1; S100P = S100 calcium binding protein P; CHI3L1 = chitinase-3-like 1; PLBD1 = phospholipase B Domain containing 1 (phospholipase B-like 1); GP1 = Glucose-6-phosphate isomerase); AZU1 = azurocidin-1; DPB = albumin D-box binding protein; SERPINA3 = serpin family A member 3 (alpha-1-antichymotrypsin); PRDX1 = peroxiredoxin-1; CAT = catalase; CST5 = cystatin D; TGM2 = transglutaminase 2.

proteome with progressive lung function impairment suggests that this approach may provide translational insights for disease stratification. That there are obvious variations in the proteome between CF patients supports this notion. Specifically, we believe that variations seen between CF subjects with similar lung function reflect limitations associated with stratification based on clinical characteristics. Agusti and colleagues have proposed that current stratification of inflammatory airway disease relies too heavily on clinical characteristics (the phenotype) and fails to appropriately account for underlying disease biology (the endotype) [39,40]. Currently, CF lung disease is classified based on characteristics such as lung function, body mass index, exacerbation frequency or sputum microbiology. That one CF subject with mild lung disease clustered amongst those with severe disease may exemplify this point. In contrast to the rest of the group, this subject was on the cusp of an exacerbation and required IV antibiotics within a month of the sputum sample being obtained. We speculate that the close proximity to a pulmonary exacerbation may have shifted the sputum proteome before changes in baseline lung function were apparent. Longitudinal measurement of the sputum proteome may offer an opportunity for the identification of biomarkers capable of capturing changes in the disease biology and recognising patients at risk of clinical deterioration before they develop symptoms or before changes can be detected in clinical characteristics, such as baseline FEV1%.

The differences in the sputum proteome in the six subjects taking

CFTR modulation therapy are intriguing, however we acknowledge that this too small a cohort to make any firm conclusions. These were limited to three subjects with class 3 or 4 CFTR mutations (Gly511Asp and Arg117His) prescribed ivacaftor, and not the other three with other mutations prescribed ivacaftor/lumacaftor. The positioning of the subjects prescribed only ivacaftor on hierarchical clustering and PCA adjacent/closer to the healthy controls rather than with those with similar lung function suggests this medication may reduce neutrophilic inflammation in subjects with these genotypes. This has not been a consistent finding in previous studies [41–43]. Of note, these findings were more marked and involved many more biological pathways when hierarchical clustering was expanded to include the entire proteome, suggesting that this approach may capture additional changes possibly related to molecular function brought about by change in CFTR function, not just due to neutrophilic inflammation. The widespread availability of ivacaftor/tezacaftor/elexacaftor (Trikafta) is timely. This potent CFTR modulator combination produces dramatic clinical improvements in CF patients homozygote or heterozygote for the Phe508del mutation. Replicating our preliminary findings in patients before and after starting ivacaftor/tezacaftor/elexacaftor is an obvious next step.

Our work has highlighted numerous sputum proteins with potential as clinical biomarkers. These could be useful in monitoring disease severity, measuring response to therapy, predicting exacerbations etc.

## a) Biological Process
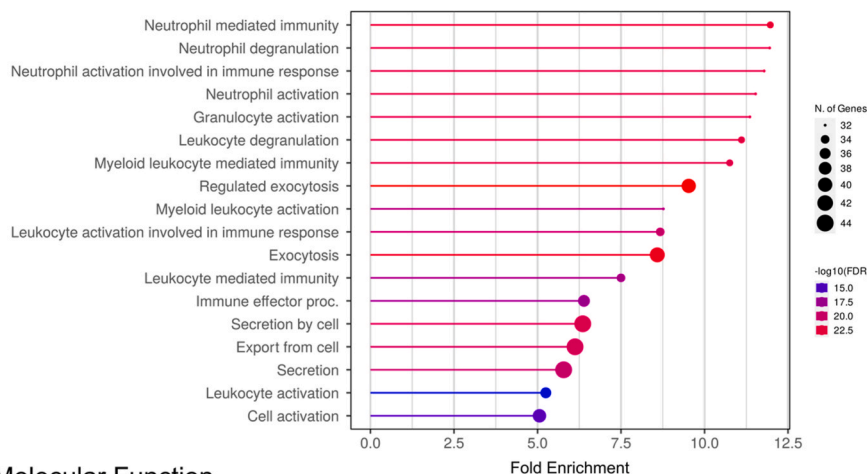


## b) Molecular Function
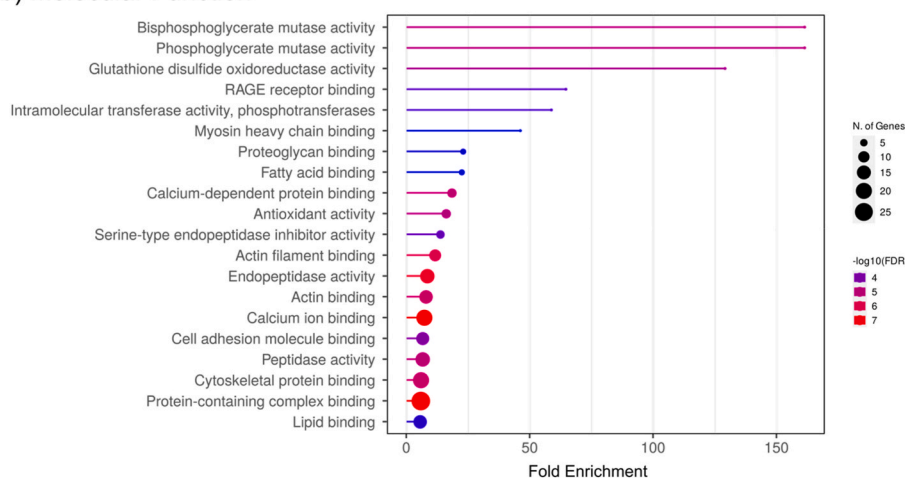


**Fig. 5. The effect of lung disease severity on molecular pathways and biological function in CF lung disease.** Proteins (87, see Table E2 in the online data supplement) that correlated ($p < 0.05$) with baseline FEV% were analysed using the ShinyGO server (http://bioinformatics.sdstate.edu/go/). The top twenty over-represented terms for (a) biological process and (b) molecular function (b) are highlighted and ranked by fold enrichment. The size of the dots represents the number of genes identified within a given term, and the colours represent statistical significance where processes highlighted in red are the most statistically significant. False discovery rate (FDR) is synonymous with adjusted p-value. Definition of abbreviations: RAGE = receptor for advanced glycation end products. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

However, it is very unlikely that a single protein biomarker will be sufficient given the complexity of the underlying biological processes in the lung and specifically the airway. There may be clinical utility in developing a carefully selected panel of sputum protein biomarkers that are able to capture as many relevant biological processes as possible, as well as inform in a range of possible clinical scenarios. To achieve this, further investigation using absolute quantification of a set of selected proteins is needed. This panel would need prospective validation in an independent cohort and then could be used to inform a predictive model for use in future studies.

There are several limitations with this study. We were not able to include a validation cohort due to the size of the study population, but the numbers are still large for this type of exploratory proteomic study. We will seek to replicate these findings within future studies. While we acknowledge a gender imbalance in both CF and healthy control cohorts, no differences were found when the sputum proteomes of male and female subjects were analysed separately (see supplementary Figure E1). Within the study, although there were a relatively small number of patients with mild disease, the overall size of the cohort compared to other proteomic studies is large and any effect was mitigated by using baseline FEV1% as a continuous variable. Lastly, the method by which sputum was collected differed between the two groups. This was unavoidable, as it is only possible to obtain sputum from healthy individuals by induction with hypertonic saline.

In conclusion, our analyses support the utility of sputum proteomics for providing insights into disease mechanisms, as well as developing potentially clinically relevant biomarkers in CF. We observed differences in the sputum proteome driven by influx and heightened activity of neutrophils with consequent disturbance of protease/antiprotease and oxidant/antioxidant balances. Of note, the proteome of CF subjects prescribed ivacaftor shifted toward the profile of healthy controls. Future work might examine the effects of CFTR modulation using sputum proteomics to provide understanding of the underlying biological effects, as well as further assist in the selection of sputum proteins for a biomarker panel to monitor lung disease.

### CRediT authorship contribution statement

**Rosemary E. Maher:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Emma Barrett:** Formal analysis, and, Writing – review & editing. **Robert J. Beynon:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition. **Victoria M. Harman:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – review & editing, Supervision. **Andrew M. Jones:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Project administration, Supervision, Funding acquisition. **Paul S.**
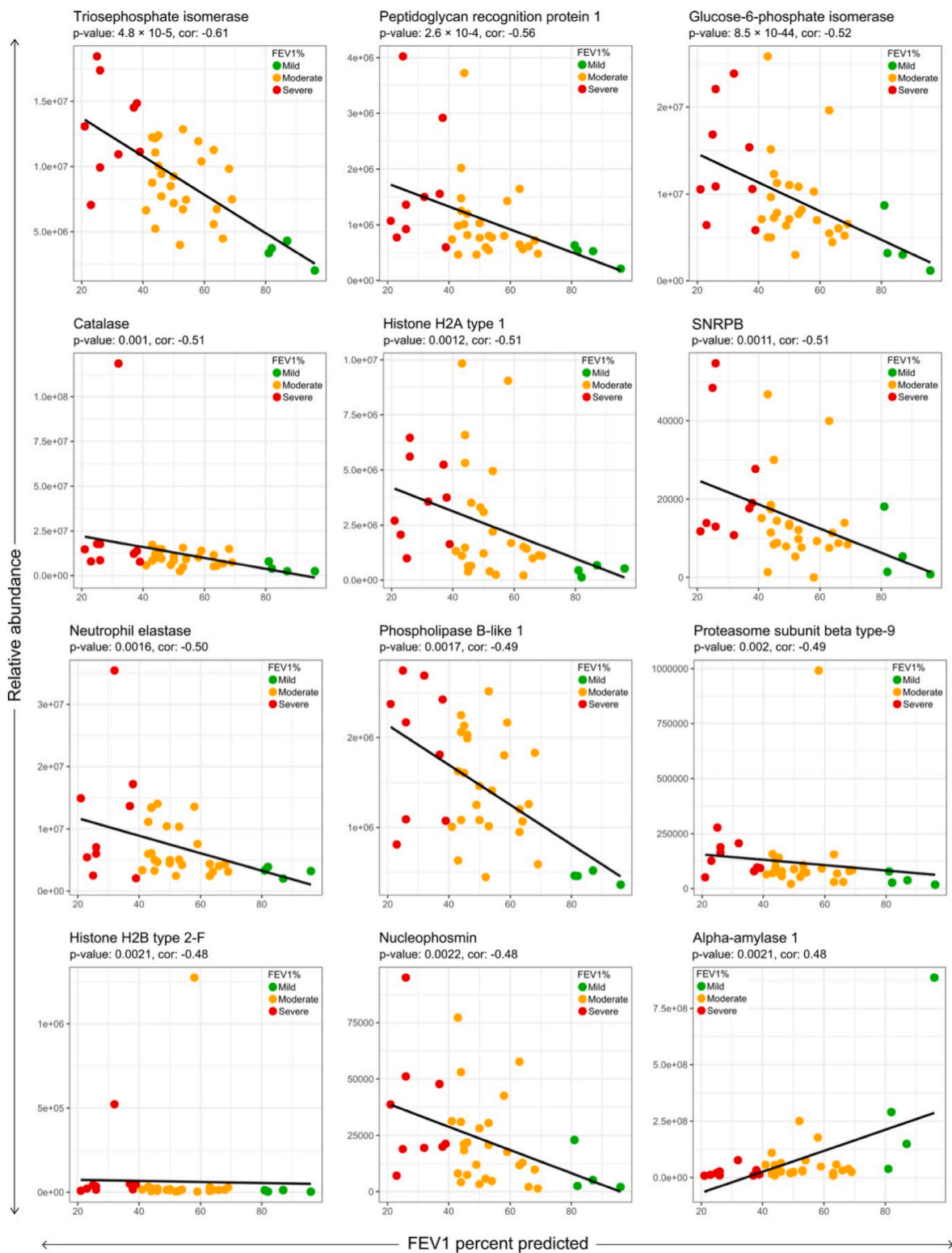
**Fig. 6. Relationship between relative abundance of selected proteins and baseline FEV1%.** For the twelve proteins with the highest confidence changes (adjusted p value) between CF and healthy controls, the abundances were plotted as a function of FEV1%. The symbols are coloured according to classification of severity, to match Fig. 3. *Definition of abbreviations*: SNRPB = small ribonuclear-associated protein B.

## Declaration of competing interest

None.

## Acknowledgements

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rmed.2022.107002.

## References

[1] M. Shteinberg, I.J. Haq, D. Polineni, J.C. Davies, Cystic fibrosis, Lancet 397 (10290) (2021) 2195–2211.

[2] D. Hartl, A. Gaggar, E. Bruscia, A. Hector, V. Marcos, A. Jung, et al., Innate immunity in cystic fibrosis lung disease, J. Cyst. Fibros. 11 (5) (2012) 363–382.

[3] C.E. Wheelock, V.M. Goss, D. Balgoma, B. Nicholas, J. Brandsma, P.J. Skipp, et al., Application of 'omics technologies to biomarker discovery in inflammatory lung diseases, Eur. Respir. J. 42 (3) (2013) 802–825.

[4] M. Roxo-Rosa, G. da Costa, T.M. Luider, B.J. Scholte, A.V. Coelho, M.D. Amaral, et al., Proteomic analysis of nasal cells from cystic fibrosis patients and non-cystic fibrosis control individuals: search for novel biomarkers of cystic fibrosis lung disease, Proteomics 6 (7) (2006) 2314–2325.

[5] I.M. Carvalho-Oliveira, N. Charro, J. Aarbiou, R.M. Buijs-Offerman, M. Wilke, T. Schettgen, et al., Proteomic analysis of naphthalene-induced airway epithelial injury and repair in a cystic fibrosis mouse model, J. Proteome Res. 8 (7) (2009) 3606–3616.

[6] E.M. DeBoer, B.D. Wagner, J. Popler, J.K. Harris, E.T. Zemanick, F.J. Accurso, et al., Novel application of aptamer proteomic analysis in cystic fibrosis bronchoalveolar lavage fluid, Proteonomics Clin. Appl. 13 (3) (2019), e1800085.

[7] S.H. Pattison, D.S. Gibson, E. Johnston, S. Peacock, K. Rivera, M.M. Tunney, et al., Proteomic profile of cystic fibrosis sputum cells in adults chronically infected with Pseudomonas aeruginosa, Eur. Respir. J. 50 (1) (2017).

[8] S.A. Gharib, T. Vaisar, M.L. Aitken, D.R. Park, J.W. Heinecke, X. Fu, Mapping the lung proteome in cystic fibrosis, J. Proteome Res. 8 (6) (2009) 3020–3028.

[9] R.D. Gray, G. MacGregor, D. Noble, M. Imrie, M. Dewar, A.C. Boyd, et al., Sputum proteomics in inflammatory and suppurative respiratory diseases, Am. J. Respir. Crit. Care Med. 178 (5) (2008) 444–452.

[10] G. Hovold, V. Palmcrantz, F. Kahn, A. Egesten, L.I. Pahlman, Heparin-binding protein in sputum as a marker of pulmonary inflammation, lung function, and bacterial load in children with cystic fibrosis, BMC Pulm. Med. 18 (1) (2018) 104.

[11] N. Mayer-Hamblett, M.L. Aitken, F.J. Accurso, R.A. Kronmal, M.W. Konstan, J.L. Burns, et al., Association between pulmonary function and sputum biomarkers in cystic fibrosis, Am. J. Respir. Crit. Care Med. 175 (8) (2007) 822–828.

[12] S.D. Sagel, R.K. Kapsner, I. Osberg, Induced sputum matrix metalloproteinase-9 correlates with lung function and airway inflammation in children with cystic fibrosis, Pediatr. Pulmonol. 39 (3) (2005) 224–232.

[13] J.S. Kim, K. Okamoto, B.K. Rubin, Pulmonary function is negatively correlated with sputum inflammatory markers and cough clearability in subjects with cystic fibrosis but not those with chronic bronchitis, Chest 129 (5) (2006) 1148–1154.

[14] A.J. Sloane, R.A. Lindner, S.S. Prasad, L.T. Sebastian, S.K. Pedersen, M. Robinson, et al., Proteomic analysis of sputum from adults and children with cystic fibrosis and from control subjects, Am. J. Respir. Crit. Care Med. 172 (11) (2005) 1416–1426.

[15] A. Hector, M.S. Kormann, I. Mack, P. Latzin, C. Casaulta, E. Kieninger, et al., The chitinase-like protein YKL-40 modulates cystic fibrosis lung disease, PLoS One 6 (9) (2011), e24399.

[16] R.D. Gray, M. Imrie, A.C. Boyd, D. Porteous, J.A. Innes, A.P. Greening, Sputum and serum calprotectin are useful biomarkers during CF exacerbation, J. Cyst. Fibros. 9 (3) (2010) 193–198.

[17] M.S. Twigg, S. Brockbank, P. Lowry, S.P. FitzGerald, C. Taggart, S. Weldon, The role of serine proteases and antiproteases in the cystic fibrosis lung, Mediat. Inflamm. 2015 (2015), 293053.

[18] P.J. Langridge, R.L. Sheehan, D.W. Denning, Microbial yield from physiotherapy assisted sputum production in respiratory outpatients, BMC Pulm. Med. 16 (2016) 23.

[19] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Stat. Soc. B 57 (1) (1995) 289–300.

[20] Z. Zhang, F. Murtagh, S. Van Poucke, S. Lin, P. Lan, Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R, Ann. Transl. Med. 5 (4) (2017) 75.

[21] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos Trans A Math Phys Eng Sci 374 (2065) (2016), 20150202.

[22] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (43) (2005) 15545–15550.

[23] S.X. Ge, D. Jung, R. Yao, ShinyGO: a graphical gene-set enrichment tool for animals and plants, Bioinformatics 36 (8) (2020) 2628–2629.

[24] S. Rahman, M. Gadjeva, Does NETosis contribute to the bacterial pathoadaptation in cystic fibrosis? Front. Immunol. 5 (2014) 378.

[25] J.H. Cho, I.P. Fraser, K. Fukase, S. Kusumoto, Y. Fujimoto, G.L. Stahl, et al., Human peptidoglycan recognition protein S is an effector of neutrophil-mediated innate immunity, Blood 106 (7) (2005) 2551–2558.

[26] A. Vargas, F. Roux-Dalvai, A. Droit, J.P. Lavoie, Neutrophil-derived exosomes: a new mechanism contributing to airway smooth muscle remodeling, Am. J. Respir. Cell Mol. Biol. 55 (3) (2016) 450–461.

[27] S. Xu, L. Zhao, A. Larsson, P. Venge, The identification of a phospholipase B precursor in human neutrophils, FEBS J. 276 (1) (2009) 175–186.

[28] S.M. Law, R.D. Gray, Neutrophil extracellular traps and the dysfunctional innate immune response of cystic fibrosis lung disease: a review, J. Inflamm. 14 (2017) 29.

[29] M.G. Surette, The cystic fibrosis lung microbiome, Ann Am Thorac Soc 11 (Suppl 1) (2014) S61–S65.

[30] D.L. Frey, S. Boutin, S.A. Dittrich, S.Y. Graeber, M. Stahl, S. Wege, et al., Relationship between airway dysbiosis, inflammation and lung function in adults with cystic fibrosis, J. Cyst. Fibros. 20(5) (2021) 754–760.

[31] G.A. Kelly-Robinson, J.A. Reihill, F.T. Lundy, L.P. McGarvey, J.C. Lockhart, G.J. Litherland, et al., The serpin superfamily and their role in the regulation and dysfunction of serine protease activity in COPD and other chronic lung diseases, Int. J. Mol. Sci. 22 (12) (2021).

[32] A.J. Holloway, J. Yu, B.P. Arulanandam, S.M. Hoskinson, T. Eaves-Pyles, Cystatins 9 and C as a novel immunotherapy treatment that protects against multidrug-resistant New Delhi metallo-beta-lactamase-1-producing Klebsiella pneumoniae, Antimicrob. Agents Chemother. 62 (3) (2018).

[33] M.H. Park, M. Jo, Y.R. Kim, C.K. Lee, J.T. Hong, Roles of peroxiredoxins in cancer, neurodegenerative diseases and inflammatory diseases, Pharmacol. Ther. 163 (2016) 1–23.

[34] J. Hui, G.K. Aulakh, S. Unniappan, B. Singh, Loss of Nucleobindin-2/Nesfatin-1 increases lipopolysaccharide-induced murine acute lung inflammation, Cell Tissue Res. 385(1) (2021) 87–103.

[35] S. Carnevali, F. Luppi, D. D'Arca, A. Caporali, M.P. Ruggieri, M.V. Vettori, et al., Clusterin decreases oxidative stress in lung fibroblasts exposed to cigarette smoke, Am. J. Respir. Crit. Care Med. 174 (4) (2006) 393–399.

[36] G. dos Santos, M.R. Rogel, M.A. Baker, J.R. Troken, D. Urich, L. Morales-Nebreda, et al., Vimentin regulates activation of the NLRP3 inflammasome, Nat. Commun. 6 (2015) 6574.

[37] S. Almuntashiri, Y. Zhu, Y. Han, X. Wang, P.R. Somanath, D. Zhang, Club cell secreted protein CC16: potential applications in prognosis and therapy for pulmonary diseases, J. Clin. Med. 9 (12) (2020).

[38] P.A. Wilmarth, M.A. Riviere, D.L. Rustvold, J.D. Lauten, T.E. Madden, L.L. David, Two-dimensional liquid chromatography study of the human whole saliva proteome, J. Proteome Res. 3 (5) (2004) 1017–1023.

[39] A. Agusti, E. Bel, M. Thomas, C. Vogelmeier, G. Brusselle, S. Holgate, et al., Treatable traits: toward precision medicine of chronic airway diseases, Eur. Respir. J. 47 (2) (2016) 410–419.

[40] R. Boaventura, O. Sibila, A. Agusti, J.D. Chalmers, Treatable traits in bronchiectasis, Eur. Respir. J. 52 (3) (2018).

[41] K.B. Hisert, S.L. Heltshe, C. Pope, P. Jorth, X. Wu, R.M. Edwards, et al., Restoring cystic fibrosis transmembrane conductance regulator function reduces airway bacteria and inflammation in people with cystic fibrosis and chronic lung infections, Am. J. Respir. Crit. Care Med. 195 (12) (2017) 1617–1628.

[42] J.K. Harris, B.D. Wagner, E.T. Zemanick, C.E. Robertson, M.J. Stevens, S.L. Heltshe, et al., Changes in airway microbiome and inflammation with ivacaftor treatment in patients with cystic fibrosis and the G551D mutation, Ann Am Thorac Soc 17 (2) (2020) 212–220.

[43] H.H. Jarosz-Griffiths, T. Scambler, C.H. Wong, S. Lara-Reyna, J. Holbrook, F. Martinon, et al., Different CFTR modulator combinations downregulate inflammation differently in cystic fibrosis, Elife 9 (2020).