



Proteomic profile of cystic fibrosis sputum cells in adults chronically infected with *Pseudomonas aeruginosa*

Sally H. Pattison¹, David S. Gibson^{1,2}, Elinor Johnston¹, Samantha Peacock³, Keith Rivera³, Michael M. Tunney⁴, Darryl J. Pappin³ and J. Stuart Elborn^{1,5}

Affiliations: ¹Centre for Experimental Medicine, Queen's University Belfast, Belfast, UK. ²Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute, University of Ulster, Londonderry, UK. ³Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. ⁴School of Pharmacy, Queen's University Belfast, Belfast, UK. ⁵National Heart and Lung Institute, Imperial College and Royal Brompton Hospital, London, UK.

Correspondence: J. Stuart Elborn, Guy Scadding Building, Imperial College London, Cae Street, London, SW3 6LY, UK. E-mail: j.elborn@imperial.ac.uk



@ERSpublications

CF sputum proteomics detects high abundance of NET proteins and identifies proteins correlating negatively with FEV₁ <http://ow.ly/hnX830aEPki>

Cite this article as: Pattison SH, Gibson DS, Johnston E, *et al.* Proteomic profile of cystic fibrosis sputum cells in adults chronically infected with *Pseudomonas aeruginosa*. *Eur Respir J* 2017; 50: 1601569 [https://doi.org/10.1183/13993003.01569-2016].

ABSTRACT Lung disease is the main cause of morbidity and mortality in cystic fibrosis (CF), and involves chronic infection and perturbed immune responses. Tissue damage is mediated mostly by extracellular proteases, but other cellular proteins may also contribute to damage through their effect on cell activities and/or release into sputum fluid by means of active secretion or cell death.

We employed MudPIT (multidimensional protein identification technology) to identify sputum cellular proteins with consistently altered abundance in adults with CF, chronically infected with *Pseudomonas aeruginosa*, compared with healthy controls. Ingenuity Pathway Analysis, Gene Ontology, protein abundance and correlation with lung function were used to infer their potential clinical significance.

Differentially abundant proteins relate to Rho family small GTPase activity, immune cell movement/activation, generation of reactive oxygen species, and dysregulation of cell death and proliferation. Compositional breakdown identified high abundance of proteins previously associated with neutrophil extracellular traps. Furthermore, negative correlations with lung function were detected for 17 proteins, many of which have previously been associated with lung injury.

These findings expand our current understanding of the mechanisms driving CF lung disease and identify sputum cellular proteins with potential for use as indicators of disease status/prognosis, stratification determinants for treatment prescription or therapeutic targets.

This article has supplementary material available from erj.ersjournals.com

Received: Aug 05 2016 | Accepted after revision: March 29 2017

Support statement: S.H. Pattison was funded by the Cystic Fibrosis Trust Project Grant PJ545. D.J. Pappin, S. Peacock and the proteomics work were part supported by the National Institutes of Health/National Cancer Institute Cancer Center Support Grant 5P30CA045508. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: None declared.

Copyright ©ERS 2017

Introduction

The lack or reduced presence at cell surfaces of fully functional cystic fibrosis transmembrane conductance regulator (CFTR) in people with CF results in a build-up of dense, dehydrated mucus that promotes infection and thus chronic inflammation [1]. Additionally, CF has been labelled a mucosal immunodeficiency syndrome because of the direct pro-inflammatory effects of CFTR dysfunction on the dysregulation of epithelial innate immunity and airway leukocytes [2]. As such, the pathophysiology of the CF lung is highly complex, involving numerous interactions between multiple human cell types and a diverse microbiota [3–5], as well as the inherent impairment of several cell types.

Proteomic studies of the CF airways have investigated protein abundance in epithelial cells cultured *in vitro* or obtained from animal models or clinical samples [6–13]. Furthermore, several studies have assessed the extracellular proteomes of CF sputum, bronchoalveolar lavage fluid (BALF) and bronchial epithelial cell secretions [14–19]. The main observations of these studies are summarised in supplementary table E1. Although these studies have all contributed to our understanding of CF, investigations of single cell types, *in vitro* cultures and animal models have limited relevance to *in situ* CF lung pathogenesis. To date, the contribution of the global proteome of all cell types in the proportions occurring naturally in CF sputum has not been investigated. We hypothesised that proteins detected in sputum cells can have a detrimental effect on CF lung health, both through their effect on cell functioning and activities, and through their release into extracellular sputum fluid by means of active secretion or cell death.

We employed quantitative comparative proteomics to characterise the global activity and protein composition of cells in CF sputum from adults chronically infected with *Pseudomonas aeruginosa* to provide insight into potential mechanisms of lung disease and to identify candidate biomarkers associated with poor lung function in this population. By employing these methods we were able to identify, in clinical samples, novel cellular proteins and activities likely to be clinically relevant as mechanisms for tissue damage and disease progression.

Methods

Study participants

12 adults with CF (mean \pm SD age 26.8 \pm 5.98 years; range 18–37 years) were recruited at Belfast City Hospital (Belfast, UK) on admission for treatment with intravenous antibiotics for an acute pulmonary exacerbation as defined by FUCHS *et al.* [20]. Seven were male and five female. All were chronically infected with *P. aeruginosa*. Their characteristics and treatment regimens are given in table 1 and supplementary table E2, respectively. Three induced sputum samples were collected from each CF participant at times of differing clinical status: untreated pulmonary exacerbation, within 24 h of completing *i.v.* antibiotic treatment of exacerbation and 4–15 weeks after completion of treatment. Lung function was measured as FEV₁ % pred prior to each sputum collection. 12 nonsmoking, healthy control participants (mean \pm SD age 29.6 \pm 4.27 years; range 25–39 years) were recruited. Eight were female and four male. A single induced sputum sample was collected from each control individual. Ethical approval for this study was granted by the Office for Research Ethics Committees Northern Ireland (09/NIR02/68) and informed consent was obtained from all subjects.

TABLE 1 Characteristics of the cystic fibrosis cohort

Subject	Age years	Sex	Pulmonary exacerbations in previous 12 months n	FEV ₁ % pred [#]	CFTR genotype
CF1	23	Female	2	35–36	F508del/F508del
CF2	36	Female	0	26–36	F508del/F508del
CF3	31	Male	1	81–102	F508del/G551D
CF4	18	Male	1	57–69	F508del/F508del
CF5	25	Male	2	52–60	F508del/F508del
CF6	37	Female	0	51–62	F508del/F508del
CF7	28	Male	0	26–30	F508del/F508del
CF11	29	Female	1	49–75	F508del/F508del
CF12	22	Male	2	31–41	F508del/F508del
CF13	19	Male	1	55–77	F508del/F508del
CF15	28	Female	1	73–98	F508del/R117H
CF16	26	Male	1	40–54	F508del/F508del

FEV₁: forced expiratory volume in 1 s; CFTR: cystic fibrosis transmembrane conductance regulator.
[#]: range measured over the three sputum collections.

Sample preparation

For each sputum sample, the cell population was harvested from mucus plugs, washed and proteins extracted. Briefly, ~1 g of mucus plug was homogenised thoroughly with 9 mL sterile phosphate-buffered saline and the cells pelleted by centrifugation. Following two further wash steps and cell straining to remove clumps, the cells were pelleted to remove all traces of supernatant. The pellet was then homogenised thoroughly in lysis buffer and subjected to repeated freeze/thaw and water bath sonication to ensure cell lysis. Insoluble debris was removed by centrifugation. All work was carried out at 4°C and in the presence of a protease inhibitor cocktail. Full method details are given in the supplementary material.

Proteomic methods

Protein digestion with trypsin and labelling with 8plex iTRAQ Reagents (AB Sciex, Warrington, UK) were performed as described previously [21] and details are given in the supplementary material. Six test samples and a reference sample were mixed for simultaneous analysis by two-dimensional liquid chromatography MudPIT (multidimensional protein identification technology) online with an LTQ Orbitrap Velos (Thermo Fisher, Waltham, MA, USA) as described previously [21]. Further details of the methods employed for MudPIT are given in the supplementary material.

Protein identification

Peaklists were generated by Mascot Distiller version 2.3 (Matrix Science, London, UK) and protein identification was performed by the Mascot search engine against the UniProt (www.uniprot.org) database plus National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) nonredundant database, taxonomy *P. aeruginosa*. Further details of the methods used in protein identification are given in the supplementary material. The mass spectrometry proteomics data have been deposited with the ProteomeXchange Consortium [22] via the PRIDE partner repository with the dataset identifier PXD001985 and 10.6019/PXD001985. Western immunoblot and ELISA were used to confirm the mass spectrometry identification and quantification data as described in the supplementary material.

Statistical analyses

SPSS Statistics (IBM, Armonk, NY, USA) was used for all statistical analysis unless otherwise stated. iTRAQ sample labels were used to calculate each protein's abundance per sample relative to the reference included in each mass spectrometry run, thus enabling comparison of protein abundance between samples. Hierarchical clustering of sample protein profiles was performed by PermutMatrix [23] using Euclidean distance dissimilarity and McQuitty's linkage criteria. Prior to cluster analysis the proteins within the dataset were tested for collinearity (bivariate Spearman correlation coefficient >0.9) and four proteins that correlated highly with others were subsequently excluded. Independent samples t-tests and Mann–Whitney U-tests were performed to identify differentially abundant proteins ($p < 0.05$) between the CF and control cohorts. Ingenuity Pathway Analysis (IPA; Ingenuity Systems, Redwood City, CA) was then employed to predict: 1) canonical pathway activity in both the CF and control consensus core proteomes (Fisher's exact test), 2) altered biological function (Bio Function) activity in CF relative to the control group (Fisher's exact test) and 3) Bio Function up- or downregulation in CF (IPA regulation z-score algorithm).

For proteins detected in >80% of CF samples, their percentage abundance within each sample was estimated by the percentage exponentially modified Protein Abundance Index (% emPAI [24]) using MS Excel (Microsoft, Redmond, WA, USA). For proteins of consistently high abundance (>1% emPAI) in CF, % emPAIs were compared between the CF and control cohorts (Mann–Whitney U-test).

Bivariate Spearman correlation analysis and univariate analysis were performed to identify proteins whose % emPAIs correlated with FEV₁ % pred and to identify independent factors. A Benjamini–Hochberg multiple comparisons adjustment was performed for the correlation analysis (MS Excel) to control the false discovery rate (FDR) at <0.05. For each of the proteins correlating negatively with FEV₁, paired t-tests and Wilcoxon signed-ranks tests were performed to compare the relative abundance levels (sample/reference) of matched CF pre-treatment and <24 h post-treatment samples to identify differential abundances associated with antibiotic treatment of exacerbation. Again, the FDR was controlled at <0.05 by a Benjamini–Hochberg multiple comparisons adjustment.

Results

Core signature proteome in CF cohort

In accordance with inclusion criteria, the CF cohort comprised adults with CF who were chronically infected with *P. aeruginosa*. The cohort's variation in FEV₁ % pred (26–102%) and collection of three sputum samples from each individual (covering clinical stability, exacerbation and following antibiotic

treatment of exacerbation) prevented proteome bias towards any particular clinical status within this population. In total, 2210 human proteins were detected including 119 which were common to all 48 samples, including those from healthy controls. A further 38 proteins were detected in all samples belonging to only one cohort, suggesting low abundance or absence in the other cohort, although differential abundance between the two cohorts cannot be proven statistically. 21 proteins were detected only in the CF cohort and 17 proteins were detected only in the control cohort. Supplementary table E3 records all proteins detected in all the samples belonging to at least one cohort.

The heatmap in figure 1, which clusters individual sputum samples according to relative abundance levels (sample/reference) of consistently detected proteins, shows that CF and control samples largely cluster separately. The two exceptions, CF3.2 and CF4.2 (marked with arrows in figure 1), are CF samples collected just after completion of antibiotic treatment for a pulmonary exacerbation.

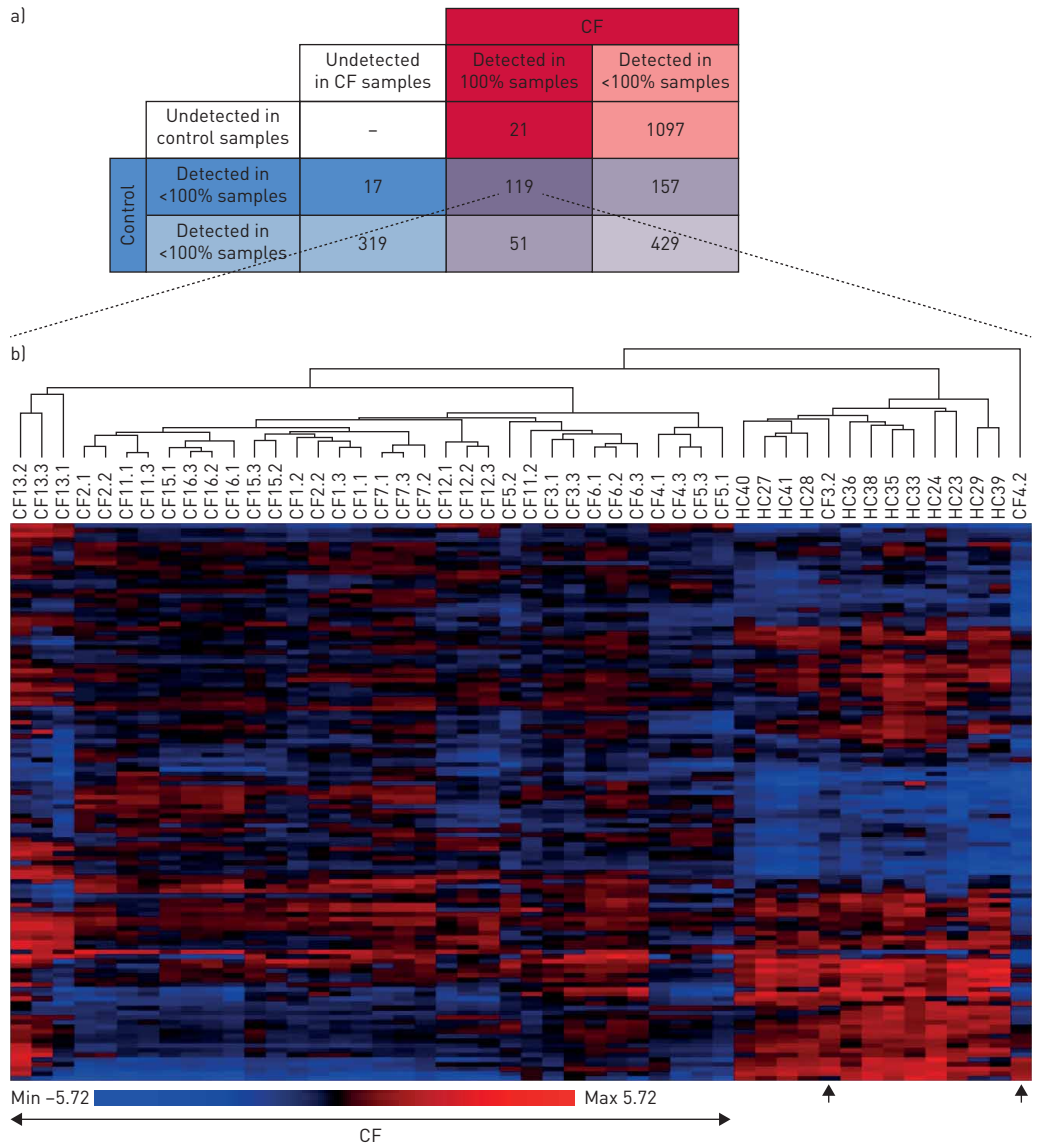


FIGURE 1 Distribution and relative abundance of human proteins detected in sputum samples collected from cystic fibrosis (CF) and healthy control cohorts. a) 119 proteins were found in all 48 samples, i.e. both CF and control, while a further 21 proteins were found exclusively in all CF samples and 17 proteins exclusively in all control samples. b) Hierarchical clustering (using Euclidean distance dissimilarity and McQuitty's linkage criteria) of individual sputum samples according to the relative abundance levels for all 119 consistently detected proteins. CF and control samples largely cluster separately, with the only exceptions being two CF samples (marked with black arrows) collected after antibiotic treatment for pulmonary exacerbation. The scale is a log₂ ratio of relative abundance.

IPA canonical pathways in CF and control cohorts

The 136 proteins detected in all control samples and the 140 proteins detected in all CF samples were treated as consensus core proteomes for the two cohorts. Qualitative individual analysis of these proteomes was performed by IPA to determine canonical pathways predicted by IPA as active in each cohort. Figure 2 shows the top five pathways predicted in the control proteome and all canonical pathways predicted by to have a greater than five-fold increase in p-value in the CF proteome compared with the control proteome, indicating likely upregulation in CF. Those predicted as most upregulated in CF are involved in immune functions, including neutrophil recruitment, rearrangement of the actin cytoskeleton, phagocytosis and T-cell signalling.

IPA Bio Function analysis of the CF differential proteome

Comparison of relative abundance levels (sample/reference) between the two cohorts (36 CF samples *versus* 12 control samples) was performed for the 119 proteins detected in all samples in order to identify cohort differences. 54 proteins were found to be less abundant and 36 more abundant in the CF *versus* control cohort (independent samples t-test and Mann–Whitney U-test, $p < 0.05$) (supplementary table E3). Abundance ratios (CF/control) for all differentially abundant proteins were included for IPA Bio Function analysis irrespective of the degree of difference. Supplementary table E4 records the Bio Functions predicted to be differentially expressed (Fisher's exact test, $p < 0.05$) in CF with regulation z-scores > 1 or < -1 (IPA regulation z-score algorithm, $p < 0.05$) and those predicted with a higher confidence level ($p < 0.0005$), but with a less noteworthy or absent regulation z-score (shaded rows in supplementary table E4).

Inflammatory Response, Cellular Movement, Cell Death, Cellular Growth and Proliferation, and Free Radical Scavenging were predicted by IPA as most altered in CF compared with the control cohort, with regard to the extent of regulation (regulation z-score) (supplementary table E4). This is largely supported by the data in supplementary table E5, which records Gene Ontology (GO) biological process classifications for the 38 proteins detected consistently in only one cohort, and the six proteins identified by Mann–Whitney U-tests as showing differential abundance between the two cohorts.

High-abundance proteins in CF sputum cells

As high-abundance proteins have greater potential for directly impacting the lung environment, we identified proteins that were consistently detected at high abundance per sample in the CF cohort and

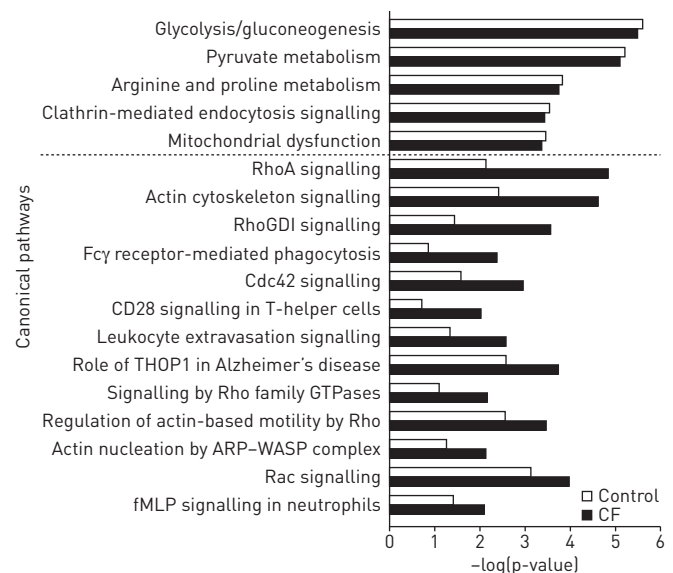


FIGURE 2 Differential activity of canonical pathways in consensus cystic fibrosis (CF) and control proteomes. Based on the set of proteins detected in all samples within a cohort, Ingenuity Pathway Analysis software applied Fisher's exact test to predict the canonical pathways likely to be active within that cohort ($p < 0.05$). The top five pathways predicted in the control cohort concern general metabolism, clathrin-mediated endocytosis signalling and mitochondrial dysfunction, and these were detected with approximately equal probabilities in both cohorts. Pathways where there is a greater than five-fold increase in the p-value in the CF proteome compared with the control proteome, indicating likely upregulation in CF, are involved in immune functions, including neutrophil recruitment, rearrangement of the actin cytoskeleton, phagocytosis and T-cell signalling. GDI: GDP dissociation inhibitor; THOP1: thimet oligopeptidase 1; ARP: actin-related protein; WASP: Wiskott–Aldrich syndrome protein; fMLP: *N*-formyl-methionyl-leucyl-phenylalanine.

determined if their abundance levels per sample differed significantly from those of the control cohort. Six proteins were found to each comprise >1% of total human cellular protein per sample (% emPAI) in >80% of CF samples. Comparison of the % emPAIs for these proteins between CF and control cohorts detected increased abundance of all six in CF (independent samples Mann–Whitney U-test, $p<0.01$). Their % emPAIs are displayed in figure 3 along with the corresponding data for the control cohort. The Human Protein Atlas [25] identifies all as being either highly or moderately expressed in bone marrow haematopoietic cells (50% of which are myelopoietic cells) and all except cathepsin G as moderately or lowly expressed in lung macrophages. No statistically significant difference in relative abundance level (sample/reference) for these six proteins was detected when comparing paired CF pulmonary exacerbation samples before and <24 h after antibiotic treatment (Wilcoxon signed-ranks test, $p<0.05$).

Sputum cellular proteins correlate with lung function

In order to identify any associations between specific proteins and lung function, the relative abundance levels (sample/reference) of the 337 proteins detected in >80% of CF samples were tested for correlation with lung function as indicated by FEV₁ % pred. Negative correlations (bivariate Spearman correlation analysis, $p<0.01$; FDR <0.05) were identified for the 17 proteins listed in table 2, implicating them as potential biomarkers of CF lung disease. Positive correlations were also identified; however, as the focus of this analysis was to identify potentially injurious proteins only the negative correlations are presented here. α_1 -Antitrypsin and olfactomedin 4 were independent factors for FEV₁ % pred (univariate analysis, $p<0.01$).

Longitudinal analysis of antibiotic treatment of CF exacerbation

Supplementary figure E1 shows that total protein yield consistently decreased for each CF participant following antibiotic treatment of exacerbation and that follow-up samples collected during clinical stability more closely resembled pre-treatment samples with respect to yield. Individual proteins identified as correlating negatively with FEV₁ % pred ($n=17$) or consistently at high abundance in CF ($n=6$) were analysed to compare the relative abundance levels (sample/reference) of matched CF pre-treatment and <24 h post-treatment samples. As shown in table 2, paired samples t-tests (including Benjamini–Hochberg multiple comparisons adjustment of FDR to <0.05) identified nine proteins with decreased abundance ($p<0.05$) following treatment of pulmonary exacerbation. Figure 1 shows that for seven of the 12 individuals with CF, samples collected during clinical stability (labelled “.3”) clustered more closely with the matched pre-treatment sample (“.1”) than the post-treatment sample (“.2”).

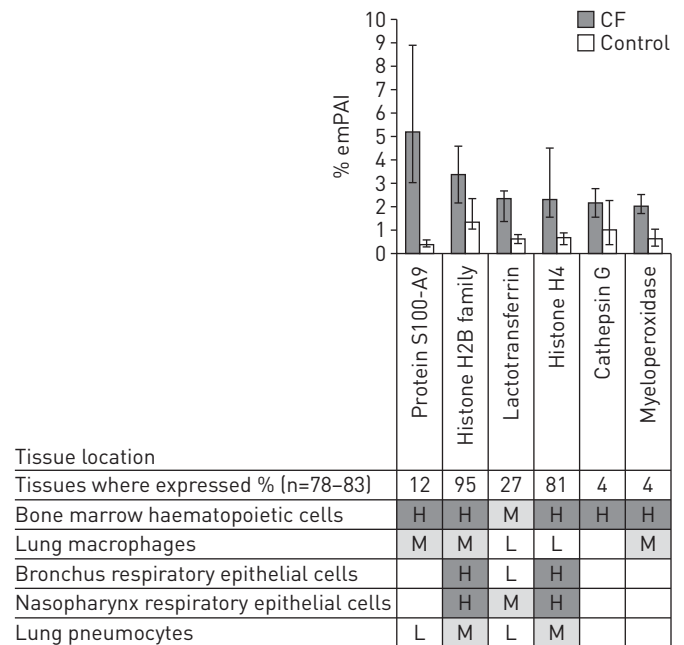


FIGURE 3 Percentage abundance estimates (median [interquartile range] percentage exponentially modified Protein Abundance Index [% emPAI]) for proteins comprising >1% of total human proteins in >80% of cystic fibrosis (CF) samples. Human Protein Atlas [25] annotations for protein expression in terms of tissue location and localised abundance are recorded. H, M and L: high, medium and low protein expression. All proteins are either highly or moderately expressed in bone marrow haematopoietic cells, 50% of which are myelopoietic cells, and all except cathepsin G are moderately or lowly expressed in lung macrophages.

TABLE 2 Correlation between relative abundance level (sample/reference) and lung function (forced expiratory volume in 1 s (FEV₁) % pred) for proteins present in >80% of cystic fibrosis (CF) samples at a false discovery rate <0.05

UniProt accession	Description	Gene symbol	Spearman correlation (FEV ₁ % pred)	p-value	Relative abundance ratio [#]
D3DSM0	Integrin β 2*	ITGB2	-0.751	0.00000	1.114
P80723	Brain acid soluble protein 1*	BASP1	-0.732	0.00000	1.343
P02675	Fibrinogen β chain**	FGB	-0.732	0.00000	1.690
P01009	α ₁ -Antitrypsin	SERPINA1	-0.674	0.00001	
P30048	Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	-0.646	0.00012	
J3QLC9	Haptoglobin*	HP	-0.617	0.00028	1.288
P08246	Neutrophil elastase	ELANE	-0.608	0.00036	
Q08722	Leukocyte surface antigen CD47*	CD47	-0.608	0.00037	1.115
Q6UX06	Olfactomedin 4**	OLFM4	-0.601	0.00011	1.620
P15144	Aminopeptidase N	ANPEP	-0.540	0.00067	
F8VV56	CD63 antigen*	CD63	-0.586	0.00068	1.338
C9JNR4	Transforming protein RhoA	RHOA	-0.578	0.00082	
P01023	α ₂ -Macroglobulin	A2M	-0.543	0.00192	
O95498	Vanin 2*	VNN2	-0.540	0.00205	1.475
O94804	Serine/threonine protein kinase 10	STK10	-0.535	0.00233	
E7ER45	Maltase-glucoamylase, intestinal**	MGAM	-0.523	0.00106	1.314
P08473	Neprilysin	MME	-0.511	0.00146	

[#]: ratio of median relative abundance levels for matched CF pre-treatment and <24 h post-treatment samples (pre-treatment/<24 h post-treatment) where protein abundance decreased (*: p<0.05; **: p<0.01) following antibiotic treatment.

ELISA validation of candidate biomarkers

As high-abundance proteins (figure 3) have greater potential for directly impacting the lung environment, we chose to validate myeloperoxidase and lactotransferrin as candidate biomarkers. Quantification of our existing samples by commercial ELISA confirmed their consistently higher abundance in CF compared with healthy control sputum cells (Mann-Whitney U-test, p<0.05) (supplementary figure E2). Additionally, this demonstrated that commercial ELISAs are a practical method for follow-up validation of candidate biomarkers in a larger independent cohort.

Discussion

In order to further elucidate the mechanisms of CF-associated tissue damage in adults with chronic respiratory *P. aeruginosa* infection, this study has for the first time specifically investigated the global proteome and functional activity of the combined population of cells present in CF sputum. As such, it uniquely studies the protein abundance of cells *in vivo*, thereby taking into account both their natural frequency in the sputum cell population and the effects of intercellular interactions. Thus, it delivers a high degree of clinical relevance to our findings. Moreover, to accurately characterise a consistent CF phenotype for adults chronically infected with *P. aeruginosa* and avoid bias due to patient individuality, we considered only those proteins detected in all samples within the CF and healthy control cohorts. Thus, our study is more stringent than previously published proteomic studies that either included proteins detected at high frequency in at least one sample [18] or used pooled samples [14–16].

In agreement with previous investigations of CF sputum or BALF [14, 15, 17, 18], we found the CF proteome to be largely distinct from that of the control cohort. IPA predicted that canonical pathways which affect or involve actin cytoskeleton rearrangement directed by the Rho family small GTPases, including RhoA, Rac1 and Cdc42, were over-represented in the CF samples. Thus, our *in vivo* data confirms previous *in vitro* observations linking CFTR deficiency with increased RhoA expression [26, 27]. This may be significant for CF lung health as impaired efferocytosis (phagocytosis of apoptotic cells) [28] and reduced expression of inducible nitric oxide synthase [27] resulting from CFTR deficiency have been reported as mediated by RhoA upregulation.

Similarly to previous proteomics studies of cell-free CF sputum/BALF, GO and IPA Bio Function analyses identified upregulation of the immune and inflammatory response [14–18]. In particular, we found increased immune cell movement and activation, and generation of reactive oxygen species. These observations fit with existing knowledge about the CF phenotype, which is typified by chronic inflammation and neutrophilia. Interestingly, an IPA predicted network identified lipopolysaccharide as a

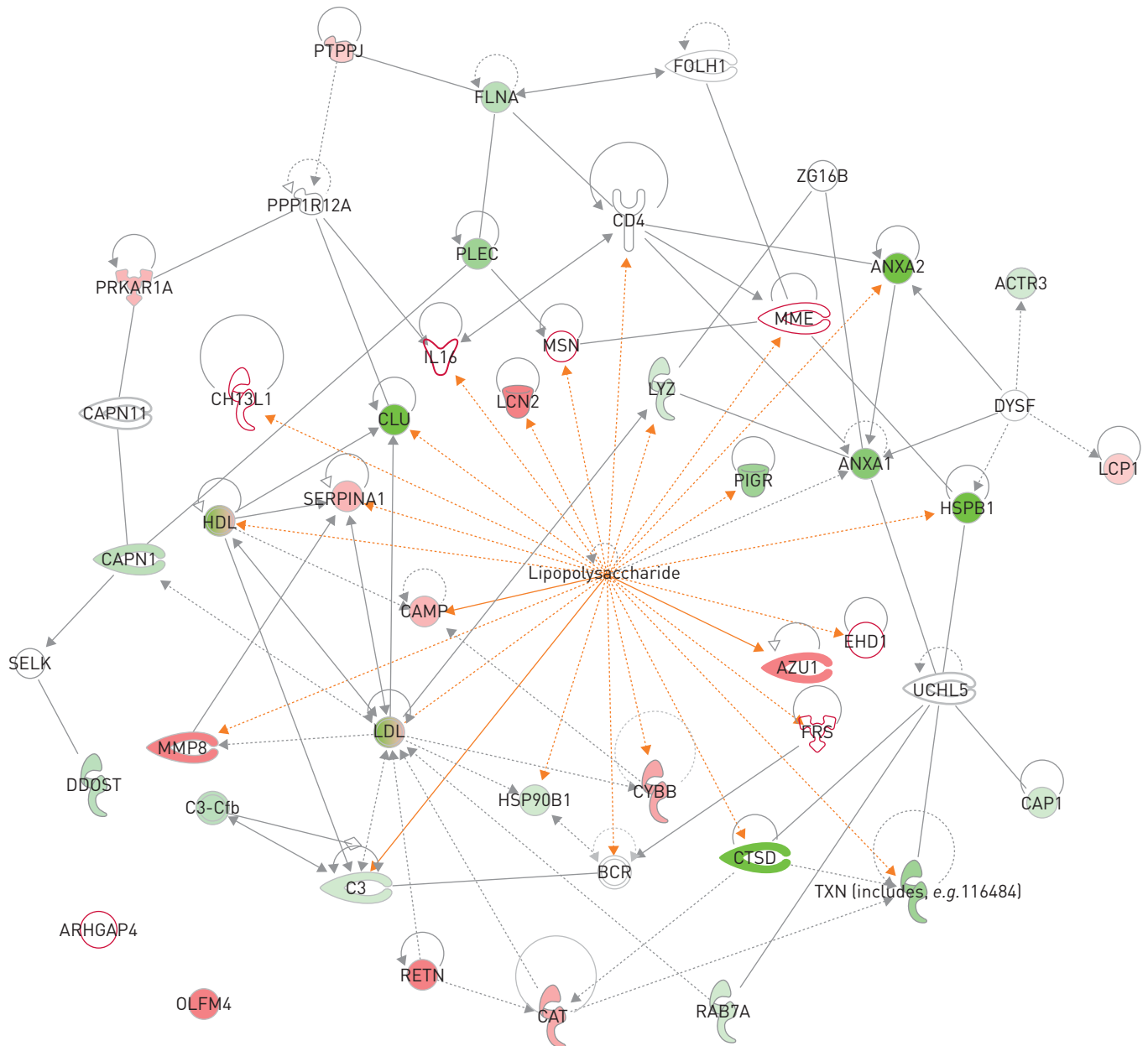


FIGURE 4 Ingenuity Pathway Analysis predicted network for proteins detected in cystic fibrosis (CF) involved in Inflammatory Response and Cell Movement Bio Functions. The intensity of red or green shading indicates the degree of increased or decreased abundance in CF relative to healthy controls, respectively. A red outline with no internal shading indicates that the protein was detected exclusively in the CF cohort. The network identifies lipopolysaccharide as a likely inflammatory trigger.

likely trigger of inflammation (figure 4). Cell death, growth and proliferation were also predicted as dysregulated in CF, in such a way that both death and proliferation of immune cells were decreased, while both were increased in nonimmune cells. These observations further support the growing understanding of the CF phenotype. Tissue damage and cell death is rampant, while excessive proliferation of respiratory epithelial cells leading to squamous cell metaplasia and dysplasia is common in CF [29, 30]. Delayed neutrophil apoptosis and clearance is thought to contribute to chronic inflammation [2], and our study detected CF-associated differential abundance of matrix metalloproteinase-8, annexin I and nicotinamide phosphoribosyltransferase, which have been implicated in delayed apoptosis and clearing of inflammatory and immune cells *in vitro* [31] and *in vivo* [32, 33].

Compositional breakdown analysis of sample proteomes found S100-A9, the histone H2B family, lactotransferrin, histone H4, cathepsin G and myeloperoxidase to be consistently present in high abundance in CF and to each be significantly more abundant in the CF cohort than the control cohort.

Importantly, all six proteins are known constituents of neutrophil extracellular trap (NET) structures, the release of which by neutrophils exerts antimicrobial activity through the degradation of virulence factors, bacterial killing and antifungal activity [34–36].

As our study focused on sputum cellular proteins we did not look specifically for NET structures, which are extracellular, and therefore do not know if they were present. However, NETs have been found previously in the CF airways fluid and the abundance of free DNA that is characteristic of NETs correlated with obstructive lung function [37]. As well as killing microbes, NETs have been found to directly induce human alveolar epithelial and endothelial cell death, with the predominant cytotoxic effect being attributed to histones and myeloperoxidase [38].

Individually, these six high-abundance proteins may also have significance for CF lung pathology and several have previously been reported as upregulated in CF sputum and/or BALF [14, 15, 18]. Importantly, differential abundance of myeloperoxidase in CF sputum fluid has been correlated with FEV₁ % pred [17], and myeloperoxidase and calgranulin B have been identified in sputum fluid as CF-associated autoantigens [16]. Cathepsin G has been implicated in degradation of surfactant protein A, so reducing the innate pulmonary antimicrobial defence [39], while S100-A9 has been shown to induce production of the gel-forming glycoprotein MUC5AC in human bronchial epithelial cells cultured *in vitro*, suggesting a potential contribution to pathological mucin hyperproduction *in vivo* [40]. In a study investigating trauma-induced lung injury, ABRAMS *et al.* [41] reported that histones induced NET formation and myeloperoxidase release *in vitro*, while histone infusion in mouse models resulted in lung structural damage and neutrophil congestion. Additionally, extracellular histones H4 and H3.3 have been implicated in asthma-related inflammation and remodelling, and in chronic obstructive pulmonary disease, respectively [42, 43]. Consequently, our consistent detection of these proteins in such high abundance in the sputum cellular fraction identifies a potential risk to lung health, either through their active secretion and NET formation or following passive release into the sputum fluid *via* cell death.

A number of proteins detected consistently in CF sputum correlated negatively with lung function measured as FEV₁ % pred. Their biological/clinical relevance and potential for causing injury are discussed more fully in the supplementary material. Longitudinal analysis of patient-matched CF samples showed differential abundances for 10 out of these 17 proteins in the CF cohort following *i.v.* antibiotic treatment of a pulmonary exacerbation. Interestingly, all the changes in abundance resulted in levels closer to those in the control cohort. Previously, SLOANE *et al.* [17] observed that for four out of 13 adults with CF, their cell-free sputum proteome profiles moved closer to those of healthy controls following antibiotic treatment of an exacerbation. However, the latter observations were made by comparison of two-dimensional gel images rather than implicating and identifying particular proteins. We also found that, although sputum total protein yield decreased in the CF cohort by completion of antibiotic treatment, it had reverted to pre-treatment levels by the clinically stable follow-up time point. This suggests that proteomic changes associated with antibiotic treatment may be temporary.

Overall, the proteins detected as negatively correlating with FEV₁ suggest that the degree of inflammation, neutrophil influx, protease activity and airway remodelling may all contribute to defining lung function. α_1 -Antitrypsin and olfactomedin 4 were identified as independent factors negatively correlating with FEV₁, although this does not demonstrate causation of harm. Acute inflammation-associated protein α_1 -antitrypsin has previously been identified as a biomarker of pulmonary disease in CF which can be predictive of response to treatment and prognosis [44]. As an inhibitor of neutrophil serine proteases, which contribute significantly to lung destruction in CF, its negative correlation with lung function might seem surprising. However, previous studies observed that plasma deficiencies during inflammation were associated with less severe lung disease in CF [45]. No association between olfactomedin 4 and CF lung health has previously been reported. However, olfactomedin 4 has been shown to downregulate pro-inflammatory responses to bacterial infection and reduce bacterial killing mediated by granule serine proteinases [46, 47], which could have the potential to negatively affect lung health.

There are a number of limitations to our study. As is typical for untargeted exploratory proteomic studies, we used highly sensitive shotgun technologies to obtain quantitative, high proteome coverage analysis of sputum cells as a novel target. The trade-off for such in-depth analysis is a relatively low sample size. Given the sample numbers it was practical to analyse using these methods, we maximised power by restricting the study to adults with CF chronically infected with *P. aeruginosa*, who experienced pulmonary exacerbation requiring hospitalisation for treatment with *i.v.* antibiotics. While CF sputum proteome responses to other microbial infections may vary, we considered this cohort most informative as 57% of the adult CF population is chronically infected with *P. aeruginosa* [48]. As this study investigated only adults the findings cannot be extended to children or individuals not chronically infected with *P. aeruginosa*. Although proteins reported as potentially damaging were detected as consistently abundant or

correlating negatively with lung function in our study, this is not proof of causation of lung injury. Additionally, the limited sample size requires that any potential significance as biomarkers either of lung function or disease severity warrants validation in a larger independent cohort. As an initial stage in this validation process we used commercial ELISAs to confirm findings in our existing cohort for myeloperoxidase and lactotransferrin. Both proteins were considered as candidate biomarkers due to their consistently high abundance in CF sputum, increased abundance in the CF compared with the control cohort and previous associations in the literature with potential for lung damage. Thus, we demonstrated that commercial ELISAs are a practical method for follow-up validation in an independent cohort. However, while cohort size and characteristics limit the scope of our study conclusions, the findings are of exploratory interest and may have future clinical value within a defined commonly encountered CF cohort. Moreover, the investigation of all types of cells in the proportions occurring naturally in sputum is novel and the approach adopted in our study could be informative for other respiratory conditions.

In summary, within the cohort studied, *i.e.* adults with chronic *P. aeruginosa* infection, we detected a distinct signature for the CF sputum cellular proteome, and have used pathway and functional analyses to predict its potential clinical impact. In particular, our data highlight the prevalence and predominance of proteins that can be associated with NET structures and have been reported as linked to tissue damage. Additionally, we identified a number of proteins that correlate negatively with lung function, many of which have previously been associated with lung injury. Overall, these findings expand the current understanding of the mechanisms behind *P. aeruginosa*-associated CF lung disease in adults and identify sputum cell proteins as candidate biomarkers for disease status, prognostic indicators or as stratification parameters for treatment prescription.

Acknowledgements

Author contributions: Conception and design: S.H. Pattison, D.J. Pappin, J.S. Elborn. Acquisition, analysis and interpretation of data: S.H. Pattison, D.S. Gibson, E. Johnston, S. Peacock, K. Rivera, M.M. Tunney, D.J. Pappin. Drafting, revising and approving manuscript: S.H. Pattison, D.S. Gibson, E. Johnston, S. Peacock, K. Rivera, M.M. Tunney, D.J. Pappin, J.S. Elborn.

References

- 1 Boucher RC. Evidence for airway surface dehydration as the initiating event in CF airway disease. *J Intern Med* 2007; 261: 5–16.
- 2 Cohen TS, Prince A. Cystic fibrosis: a mucosal immunodeficiency syndrome. *Nat Med* 2012; 18: 509–519.
- 3 Cox MJ, Allgaier M, Taylor B, *et al.* Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS One* 2010; 5: e11044.
- 4 Guss AM, Roeselers G, Newton IL, *et al.* Phylogenetic and metabolic diversity of bacteria associated with cystic fibrosis. *ISME J* 2011; 5: 20–29.
- 5 Delhaes L, Monchy S, Fréalle E, *et al.* The airway microbiota in cystic fibrosis: a complex fungal and bacterial community – implications for therapeutic management. *PLoS One* 2012; 7: e36313.
- 6 Bensalem N, Ventura AP, Vallée B, *et al.* Down-regulation of the anti-inflammatory protein annexin A1 in cystic fibrosis knock-out mice and patients. *Mol Cell Proteomics* 2005; 4: 1591–1601.
- 7 Roxo-Rosa M, da Costa G, Luidert TM, *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* 2006; 6: 2314–2325.
- 8 Carvalho-Oliveira I, Charro N, Aarbiou J, *et al.* Proteomic analysis of naphthalene-induced airway epithelial injury and repair in a cystic fibrosis mouse model. *J Proteome Res* 2009; 8: 3606–3616.
- 9 Gomes-Alves P, Imrie M, Gray RD, *et al.* SELDI-TOF biomarker signatures for cystic fibrosis, asthma and chronic obstructive pulmonary disease. *Clin Biochem* 2010; 43: 168–177.
- 10 Ciavardelli D, D'Orazio M, Pieroni L, *et al.* Proteomic and ionomic profiling reveals significant alterations of protein expression and calcium homeostasis in cystic fibrosis cells. *Mol Biosyst* 2013; 9: 1117–1126.
- 11 Ruaniyar N, Gupta V, Balch WE, *et al.* Quantitative proteomic profiling reveals differentially regulated protein in cystic fibrosis cells. *J Proteome Res* 2014; 13: 4668–4675.
- 12 Lavallée-Adam M, Ruaniyar N, McClatchy DB, *et al.* PSEA-Quant: a protein set enrichment analysis on label-free and label-based protein quantification data. *J Proteome Res* 2014; 13: 5496–5509.
- 13 Jeanson L, Guerrero IC, Papon JF, *et al.* Proteomic analysis of nasal epithelial cells from cystic fibrosis patients. *PLoS One* 2014; 9: e108671.
- 14 Gray RD, MacGregor G, Noble D, *et al.* Sputum proteomics in inflammatory and suppurative respiratory diseases. *Am J Respir Crit Care Med* 2008; 178: 444–452.
- 15 MacGregor G, Gray RD, Hilliard TN, *et al.* Biomarkers for cystic fibrosis lung disease: application of SELDI-TOF mass spectrometry to BAL fluid. *J Cyst Fibros* 2008; 7: 352–358.
- 16 Pedersen SK, Sloane AJ, Prasad SS, *et al.* An immunoproteomic approach for identification of clinical biomarkers for monitoring disease. *Mol Cell Proteomics* 2005; 4: 1052–1060.
- 17 Sloane AJ, Lindner RA, Prasad SS, *et al.* Proteomic analysis of sputum from adults and children with cystic fibrosis and from control subjects. *Am J Respir Crit Care Med* 2005; 172: 1416–1426.
- 18 Gharib S, Vaisar T, Aitken M, *et al.* Mapping the lung proteome in cystic fibrosis. *J Proteome Res* 2009; 8: 3020–3029.
- 19 Peters-Hall JR, Brown KJ, Pillai DK, *et al.* Quantitative proteomics reveals an altered cystic fibrosis *in vitro* bronchial epithelial secretome. *Am J Respir Cell Mol Biol* 2015; 53: 22–32.

- 20 Fuchs HJ, Borowitz DS, Christiansen DH, *et al.* Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *N Engl J Med* 1994; 331: 637–642.
- 21 Scuoppo C, Miething C, Lindqvist L, *et al.* A tumour suppressor network relying on the polyamine–hypusine axis. *Nature* 2012; 487: 244–248.
- 22 Vizcaino JA, Deutsch EW, Wang R, *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* 2014; 32: 223–226.
- 23 Caraux G, Pinloche S. PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics* 2005; 21: 1280–1281.
- 24 Ishihama Y, Oda Y, Tabata T, *et al.* Exponentially modified Protein Abundance Index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 2005; 4: 1265–1272.
- 25 Uhlen M, Bjorling E, Agaton C, *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* 2005; 4: 1920–1932.
- 26 Perez A, Issler AC, Cotton CU, *et al.* CFTR inhibition mimics the cystic fibrosis inflammatory profile. *Am J Physiol Lung Cell Mol Physiol* 2007; 292: L383–L395.
- 27 Kreiselmeier NE, Kraynack NC, Corey DA, *et al.* Statin-mediated correction of STAT1 signaling and inducible nitric oxide synthase expression in cystic fibrosis epithelial cells. *Am J Physiol Lung Cell Mol Physiol* 2003; 285: L1286–L1295.
- 28 Vandivier RW, Richens TR, Horstmann SA, *et al.* Dysfunctional cystic fibrosis transmembrane conductance regulator inhibits phagocytosis of apoptotic cells with proinflammatory consequences. *Am J Physiol Lung Cell Mol Physiol* 2009; 297: L677–L686.
- 29 Voynow JA, Fischer BM, Roberts BC, *et al.* Basal-like cells constitute the proliferating cell population in cystic fibrosis airways. *Am J Respir Crit Care Med* 2005; 172: 1013–1018.
- 30 Hajj R, Lesimple P, Nawrocki-Raby B, *et al.* Human airway surface epithelial regeneration is delayed and abnormal in cystic fibrosis. *J Pathol* 2007; 211: 340–350.
- 31 Scannell M, Flanagan MB, deStefani A, *et al.* Annexin-1 and peptide derivatives are released by apoptotic cells and stimulate phagocytosis of apoptotic neutrophils by macrophages. *J Immunol* 2007; 178: 4595–4605.
- 32 Jia SH, Li Y, Parodo J, *et al.* Pre-B cell colony-enhancing factor inhibits neutrophil apoptosis in experimental inflammation and clinical sepsis. *J Clin Invest* 2004; 113: 1318–1327.
- 33 Gueders MM, Balbin M, Rocks N, *et al.* Matrix metalloproteinase-8 deficiency promotes granulocytic allergen-induced airway inflammation. *J Immunol* 2005; 175: 2589–2597.
- 34 Brinkmann V, Reichard U, Goosmann C, *et al.* Neutrophil extracellular traps kill bacteria. *Science* 2004; 303: 1532–1535.
- 35 Urban CF, Ermert D, Schmid M, *et al.* Neutrophil extracellular traps contain calprotectin, a cytosolic protein complex involved in host defense against *Candida albicans*. *PLoS Pathog* 2009; 5: e1000639.
- 36 Bianchi M, Niemiec MJ, Siler U, *et al.* Restoration of anti-*Aspergillus* defense by neutrophil extracellular traps in human chronic granulomatous disease after gene therapy is calprotectin-dependent. *J Allergy Clin Immunol* 2011; 127: 1243–1252.
- 37 Marcos V, Zhou Z, Yildirim AO, *et al.* Free DNA in cystic fibrosis airway fluids correlates with airflow obstruction. *Mediators Inflamm* 2015; 2015: 408935.
- 38 Saffarzadeh M, Juenemann C, Queisser MA, *et al.* Neutrophil extracellular traps directly induce epithelial and endothelial cell death: a predominant role of histones. *PLoS One* 2012; 7: e32366.
- 39 Rubio F, Cooley J, Accurso FJ, *et al.* Linkage of neutrophil serine proteases and decreased surfactant protein-A (SP-A) levels in inflammatory lung disease. *Thorax* 2004; 59: 318–323.
- 40 Kang JH, Hwang SM, Chung IY. S100A8, S100A9 and S100A12 activate airway epithelial cells to produce MUC5AC via extracellular signal-regulated kinase and nuclear factor- κ B pathways. *Immunology* 2015; 144: 79–90.
- 41 Abrams ST, Zhang N, Manson J, *et al.* Circulating histones are mediators of trauma-associated lung injury. *Am J Respir Crit Care Med* 2013; 187: 160–169.
- 42 Calvo FQ, Fillet M, de Seny D, *et al.* Biomarker discovery in asthma-related inflammation and remodeling. *Proteomics* 2009; 9: 2163–2170.
- 43 Barrero CA, Perez-Leal O, Aksoy M, *et al.* Histone 3.3 participates in a self-sustaining cascade of apoptosis that contributes to the progression of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013; 188: 673–683.
- 44 Sagel SD, Thompson V, Chmiel JF, *et al.* Effect of treatment of cystic fibrosis pulmonary exacerbations on systemic inflammation. *Ann Am Thorac Soc* 2015; 12: 708–717.
- 45 Mahadeva R, Westerbeek R, Perry D, *et al.* α_1 -Antitrypsin deficiency alleles and the *Taq-I* G→A allele in cystic fibrosis lung disease. *Eur Respir J* 1998; 11: 873–879.
- 46 Liu W, Yan M, Liu Y, *et al.* Olfactomedin 4 down-regulates innate immunity against *Helicobacter pylori* infection. *Proc Natl Acad Sci USA* 2010; 107: 11056–11061.
- 47 Liu W, Yan M, Liu Y, *et al.* Olfactomedin 4 inhibits cathepsin C-mediated protease activities, thereby modulating neutrophil killing of *Staphylococcus aureus* and *Escherichia coli* in mice. *J Immunol* 2012; 189: 2460–2467.
- 48 Cystic Fibrosis Foundation. Patient Registry: 2015 Annual Data Report to the Center Directors. Bethesda, Cystic Fibrosis Foundation, 2016.