Large-scale label-free quantitative mapping of the sputum proteome

Dominic Burg[1, 2†], James P R Schofield[1, 2†*], Joost Brandsma[2], Doroteya Staykova[1], Caterina Folisi[1], Aruna Bansal[3], Ben Nicholas[2], Yang Xian[4], Anthony Rowe[5], Julie Corfield[6], Susan Wilson[2], Jonathan Ward[2], Rene Lutter[7, 8], Louise Fleming[9], Dominick E Shaw[10], Per S Bakke[11], Massimo Caruso[12], Sven-Erik Dahlen[13], Stephen J. Fowler[14], Simone Hashimoto[15], Ildikó Horváth[16], Peter Howarth[2], Norbert Krug[17], Paolo Montuschi[18], Marek Sanak[19], Thomas Sandström[20], Florian Singer[21], Kai Sun[4], Ioannis Pandis[4], Charles Auffray[22], Ana R Sousa[23], Ian M Adcock[24], Kian Fan Chung[9], Peter J Sterk[7], Ratko Djukanović[2, #], Paul J Skipp[1, #] and the U-BIOPRED Study Group[25]

[1]Centre for Proteomic Research, University of Southampton, UK

[2]NIHR Southampton Respiratory Biomedical research unit, University Hospital Southampton, UK

[3]Acclarogen Ltd, Cambridge, UK

[4]Data Science Institute, Imperial College, London, UK

[5]Janssen Research & Development, Buckinghamshire, UK

[6]Areteva Ltd, Nottingham, UK

[7]AMC, Department of Experimental Immunology, University of Amsterdam, Amsterdam, The Netherlands

[8]AMC, Department of Respiratory Medicine, University of Amsterdam, Amsterdam, The Netherlands

[9]Airways Disease, National Heart and Lung Institute, Imperial College, London & Royal Brompton NIHR Biomedical Research Unit, London, United Kingdom

[10]Respiratory Research Unit, University of Nottingham, UK

[11]Institute of Medicine, University of Bergen, Bergen, Norway

[12]Dept. of Clinical and Experimental Medicine Hospital University, University of Catania, Catania, Italy.

[13]The Centre for Allergy Research, The Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

[14]Respiratory and Allergy Research Group, University of Manchester, Manchester, UK.

[15]Dept. of Respiratory Medicine, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands.

[16]Dept. of Pulmonology, Semmelweis University, Budapest, Hungary

[17]Fraunhofer Institute for Toxicology and Experimental Medicine Hannover, Hannover, Germany.

[18]Faculty of Medicine, Catholic University of the Sacred Heart, Rome, Italy.

[19]Laboratory of Molecular Biology and Clinical Genetics, Medical College, Jagiellonian University, Krakow, Poland

[20]Dept. of Medicine, Dept of Public Health and Clinical Medicine Respiratory Medicine Unit, Umeå University, Umeå, Sweden.

[21]University Children's Hospital Zurich, Zurich, Switzerland.

[22]European Institute for Systems Biology and Medicine, CNRS-ENS-UCBL-INSERM, Université de Lyon, France

[23]Respiratory Therapeutic Unit, GSK, Stockley Park, UK.

[24]Cell and Molecular Biology Group, Airways Disease Section, National Heart and Lung Institute, Imperial College London, Dovehouse Street, London, UK

[25]A full list of the U-BIOPRED Study Group members and their affiliations can be found in the acknowledgements.

*To whom correspondence should be addressed: James P R Schofield at J.P.R.Schofield@soton.ac.uk, +44 (0) 23 80594204
†Equally contributing joint first authors
Joint senior authors

Abbreviations:

U-BIOPRED - Unbiased BIOmarkers Predictive of REspiratory Disease outcomes

HDMSE – High definition mass spectrometry

COPD – Chronic Obstructive Pulmonary Disease

CE - collision energy

BAL - bronchoalveolar lavage

Abstract: Analysis of induced sputum supernatant is a minimally invasive approach to study the epithelial lining fluid and, thereby, provide insight into normal lung biology and the pathobiology of lung diseases. We present here a novel proteomics approach to sputum analysis developed within the U-BIOPRED (Unbiased BIOmarkers Predictive of REspiratory Disease outcomes) international project. We present practical and analytical techniques to optimise the detection of robust biomarkers in proteomic studies. The normal sputum proteome was derived using data-independent HDMS^E applied to 40 healthy non-smoking participants, which provides an essential baseline from which to compare modulation of protein expression in respiratory diseases. The "core" sputum proteome (proteins detected in ≥40 % of participants) was composed of 284 proteins and the extended proteome (proteins detected in ≥3 participants) contained 1666 proteins. Quality control procedures were developed to optimise the accuracy and consistency of measurement of sputum proteins and analyse the distribution of sputum proteins in the healthy population. The analysis showed that quantitation of proteins by HDMS^E is influenced by several factors, with some proteins being measured in all participants' samples and with low measurement variance between samples from the same patient. The measurement of some proteins is highly variable between repeat analyses, susceptible to sample processing effects, or difficult to accurately quantify by mass spectrometry. Other proteins show high inter-individual variance. We also highlight that the sputum proteome of healthy individuals is related to sputum neutrophil levels, but not gender or allergic sensitisation. We illustrate the importance of design and interpretation of disease biomarker studies considering such protein population and technical measurement variance.

Key words: asthma, proteomics, biomarkers, U-BIOPRED, sputum, HDMS^E, unbiased, variance, allergic, neutrophil

INTRODUCTION

Sputum induction is a widely applied method of sampling the epithelial lining fluid that lines the lower airways constituting the tracheobronchial tree. It enables assessment of both the cellular and extracellular environments in the lung[1-3] and is particularly useful in the study of inflammatory respiratory diseases, providing insight into the immune and structural cell populations and their secreted products. Initial studies of induced sputum focused on inflammatory cell counts and targeted quantification of soluble proteins by enzyme-linked immunosorbent assay (ELISA). Such analyses identified several induced sputum biomarkers as valuable in the description of inflammation in common chronic airway diseases including asthma and chronic obstructive pulmonary disease (COPD)[8,9], providing insight into determinants of disease severity[9a] and relevant pathophysiological abnormalities, such as airway hyperresponsiveness [10] and changes in airway geometry [11]. Combined with studies of cell function such as chemotactic activity; quantification of cytokines and chemokines in the sputum fluid phase has provided a better understanding of the extent to which individual mediators contribute to inflammation, thereby providing initial stratification of respiratory disease.

Methods for global, unbiased analysis, that do not select *a priori* which analytes are measured, including transcriptomics, proteomics and lipidomics, appear useful for stratifying disease[12, 13]. However, only a limited number of unbiased proteomic studies focusing on the lungs have been published to date. Ten years ago, we described the first sputum proteome, applying a shotgun method to an induced sputum sample from a female smoker with no detectable evidence of lung disease[2]. Since then, there have been a number of reports of this approach in COPD[18, 19] including a study highlighting the utility of protein network analysis in sputum[20], and a large study combining proteomic and transcriptomic analyses[21]. Likewise, limited studies of sputum have been performed to study asthma and some have been relatively low throughput[22, 23]. Apart from one study, by Titz *et al.*[21], coverage of the sputum proteome remains low. Our previous study highlighted some overlap between sputum proteomes and proteomes of other sample types, namely bronchoalveolar lavage (BAL) and saliva. In recent years, attention has been drawn to the repeated failure of published biomarkers to translate to the clinic[25-29]. Such failure is often attributable to study design and validation, insufficient sample size and inappropriate experimental methodology. Problems with sample size are beginning to be addressed in proteomics[30], particularly with the advent of data independent approaches such as MS^E [31, 32] which allow absolute comparison of samples without the inherent limitations associated with multiplexing, labelling or spectral counting-based quantitation, [33-35]. Despite the utility of MS^E for large clinical studies, there is very little information on the effect of measuring samples over extended periods and resulting data variability.

The approach to quality control in the analysis of human BAL samples using repeated measurements and pooled samples published by the Moseley group at Duke University [36, 37] is a standout example of the necessary approach required for clinical studies; however, sample sizes in these studies have been relatively small.

In the current study, we have applied state of the art quantitative HDMS^E analysis to a large set of sputum samples to advance on the sputum proteome previously reported[2]. As part of the method evaluation, we explored the impact of granulocytic infiltration of the airways, participant gender and other common demographics on the sputum proteome. Given the high prevalence of allergic sensitisation to common airborne allergens (e.g. house dust mite and pollens) in the general population, we also examined how atopy, defined by sensitisation to at least one common aero-allergen, affects the sputum proteome. As a key component of the study, we assessed variability in proteomic measurements and considered the impact of such variability on biomarker discovery. Using repeated measurements, pooled samples, comparison between individuals and to serum samples from the same study participants, we assessed the likely source of variability in measurements on a protein by protein basis. We discuss the impact of variability on effective sample size and statistical power for comparative studies. Finally, we have performed an in-depth analysis of tissue and cellular origins of proteins from previous proteomic studies and defined the accessible functional proteomic space using functional enrichment analysis.

MATERIALS AND METHODS

Study design and participant characteristics

The U-BIOPRED study was performed in 14 European clinical centres with extensive experience in sputum induction and processing. The clinical study has been described elsewhere[39] and the protocol was approved by all local Ethics Review Boards. Participants gave their written informed consent for extensive characterisation using routine clinical protocols, including, lung function tests, assessment of sensitization to common aero-allergens, and haematological and biochemistry blood tests (reported in detail in Shaw *et al.,* 2015[38]). Samples were stored in a central biobank (CIGMR Biobank, University of Manchester) where they were blinded. Identity of the samples were un-blinded only after all the mass spectrometric analyses and data pre-processing had been completed.

Table 1 Demographics and sputum cell characteristics

	N	40
Age (mean ± SEM)		**35.4** ± 2.3
Weight (kg) (mean ± SEM)		**85** ± 2.14
BMI (mean ± SEM)		**25.62** ± 0.51
Gender (F/M)		**29/11**
Race (% white Caucasian)		**90%**
Smoking history ex smokers (count, % of total)		**5** (12.5%)
Atopy positive (count, % of total)		**13** (32.5%)
Positive IgE Assay (count, % of total)		**7** (17.5%)
Positive Skin Prick Assay (count, % of total)		**12** (30.0%)
FEV1/FVC predicted % (mean ± SEM)		**83.58** ± 0.38
Sputum % Neutrophil (median, range)		**37.95** (2.71 - 88.34)
Sputum % Eosinophil (median, range)		**0.00** (0.00 - 2.57)
Sputum % Lymphocytes (median, range)		**1.22** (0.00 - 7.76)
Sputum % Macrophages (median, range)		**60.30** (7.11 - 96.10)
Sputum % Squamous epithelial (median, range)		**14.70** (0.00 - 39.20)

Forty healthy individuals (mean age 36.9 years, range 18 to 65, 70% male), provided sputum samples considered representative of the bronchial compartment, i.e. ≤ 40% contaminating squamous cells (Table 1). The frequency of atopy, demonstrated by positive skin or serum IgE specific for at least one common aero-allergen test, was 32.5%. The atopic and non-atopic participants did not differ in respect of sputum cell counts, including sputum eosinophils (<2% of total inflammatory cells in all participants), blood eosinophils, and lung function. As expected, total serum IgE concentrations were higher in atopic individuals ($p<5E^{-5}$), but, surprisingly, serum LDH was also slightly higher in the atopic participants (p=0.01) while alkaline phosphatase was lower (p=0.01).

Sputum induction and processing

Sputum induction with nebulised hypertonic saline (4.5% NaCl) and sample processing were performed in accordance with the recommendations of the European Respiratory Society Task Force on induced sputum methods[39]. Uniformity of methods was ensured by all study sites using standard operating procedures (SOP) and centralised training. For consistency required for comparison with proteomes in patients with disease (asthma or COPD), all participants were pre-medicated with the β2-agonist, salbutamol, given as standard, to prevent excessive bronchoconstriction in patients with airways disease.

Induced Mucoid portions of the induced sputum were selected with forceps to reduce salivary contamination,

weighed and solubilised at room temperature with 6.8mM dithioerythritol (DTE) in HEPES buffered saline, added at a 4:1 w/v ratio. The solution was filtered through a 100-μm filter, centrifuged at 400 x g to remove the cell pellet, further centrifuged at 12,000 x g to remove cell debris, both at 4°C, and stored at -80°C. The cell pellets were processed for quantification of alive or dead respiratory cells, squamous cells and differential inflammatory cell counts (by Diff-Quick rapid Romanowsky stain); eosinophil, neutrophil, macrophage/monocyte, lymphocyte and mast cell/basophil counts were reported as percentages of total inflammatory cells, while squamous cells were reported as a percentage of total cell counts.

Protein isolation and preparation for analysis

Sputum samples were thawed to room temperature before taking 100-μl aliquots for extraction of lipids using a semi-automated Bligh-Dyer protocol (Bligh and Dyer, 1959) on a robotic liquid handling platform (Freedom EVO 100; TECAN, Männedorf, Switzerland). Briefly, each sample was made up to a volume of 800 μl with 0.9% saline solution before adding 2 ml of methanol (MeOH) and 1 ml of dichloromethane (DCM) and 10 μl of anti-oxidant (5 mg ml^{-1} butylated hydroxytoluene in MeOH). Samples were centrifuged at 1000 x g for 10 min at 10°C to produce protein pellets which were snap-frozen in liquid nitrogen and stored at -80°C.

In preparation for analysis, the frozen protein pellets were thawed to room temperature, dissolved in 150μl of 50% trifluoroethanol, 50mM ammonium bicarbonate and heated at 60°C for 30 min. A pool for quality control was prepared with equal protein amounts from 40 different sputum samples (including healthy participants and participants with a diagnosis of asthma). Pool samples were processed and analysed in parallel in batches containing 11 analytical samples and one pool. Dissolved protein pellets were reduced, alkylated and digested with trypsin. Peptide samples were filtered using a 10KDa cut-off ultrafiltration device (Millipore) and the filtrate lyophilised *in-vacuo*. Samples were dissolved in 3% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA) in preparation for reverse phase clean-up, performed according to the manufacturer's instructions using C18 spin tips (Protea Biosciences). Following elution, peptides were lyophilised and stored at -80°C prior to analysis.

Serum collection and processing

Clotted venous blood samples were centrifuged at 1,000 x g for 10 min, and collected supernatants stored at -80°C. A pooled serum sample was created as for sputum. In order to increase the number of identifications, the 12 most abundant proteins were immunodepleted using disposable agarose columns (Pierce / Thermo-Fisher) and eluates reduced, alkylated, digested and lyophilised. Peptide extracts were then re-suspended in 3% ACN, 0.1% TFA and desalted using 96 well RP solid phase extraction plates

(3M Empore). Eluates were transferred to separate microcentrifuge tubes, lyophilised and stored on ice until mass spectrometry.

Mass spectrometry

Peptide extracts were re-suspended in buffer A, (3% ACN, 0.1% Formic acid (*v/v*) and the concentration measured using a Direct Detect System (Millipore). An internal standard mixture of *E. coli* ClpB Hi3 standard (Waters), yeast enolase (ENO) and yeast alcohol dehydrogenase (ADH) was added to a final concentration in 20μl of 250ng/μl sputum peptide, 12.5 fmol/μl ClpB, 12.5 fmol/μl ENO, and 8.75 fmol/μl ADH (serum was 25% more concentrated).

Samples were analysed in duplicate, sequentially (not spread across batches), via HDMS^E on a Waters Synapt G2S high definition mass spectrometer coupled to a nanoAcquity UPLC system. 4μl of peptide extract was injected onto a C18 BEH trapping column (Waters) and washed with buffer A for 5 min at 5μl/min. Peptides were separated using a 25cm T3 HSS C18 analytical column (Waters) with a linear gradient of 3-50% ACN + 0.1% formic acid over 50 minutes at a flow rate of 0.3μl/min. Eluted samples were sprayed directly into the mass spectrometer operating in MS^E mode. Data were acquired from 50 to 2000 *m/z* with the quadrupole in RF mode using alternate low and elevated collision energy (CE) scans, resolution of 35,000. Low CE was 5V and elevated CE ramp from 15 to 40V. Ion mobility separation was implemented prior to fragmentation using a wave velocity of 650 m/s and wave height of 40V. The lock mass Glu-fibrinopeptide, (M+2H)$^{+2}$, *m/z* = 785.8426) was infused at a concentration of 100 fmol/μl at a flow rate of 250 nl/min and acquired every 60 sec.

Database searching and curation

Raw data were processed using a custom package (Regression tester) based upon executable files from ProteinLynx Global Server 3.0 (Waters). The optimal setting for peak detection across the dataset was determined using Threshold inspector (Waters) and these thresholds were applied: low energy = 100 counts; high energy = 30 (for serum this was set to 25) and a total energy count threshold of 750. Database searches were performed using regression tester and searched against the Uniprot human reference database (20/11/2014; 20,229 entries) with added sequence information for internal standards. A maximum of two missed cleavages was allowed for tryptic digestion and the variable modification was set to contain oxidation of methionine and carboxyamidomethylation of cysteine. Precursor and product ion mass tolerances were calculated automatically during data processing and the false discovery rate (FDR) was set at 4%. We report only proteins identified in at least two patient samples, which results in a FDR below 1%[40]. Only proteins identified in each technical replicate of at least two patient samples were considered; thus, the false positive rate is minimized, since chemical noise is

random in nature and does not replicate across injections. Quantity was estimated in absolute amounts using the Top 3 method[32, 41]. The ion accounting output files[42] were compiled and summary information generated from search log files using custom Python scripts. Information contained in ion accounting files were collated into a single .csv document using a custom Python script.

Data filtering and normalisation

Protein identifications collated from the ion accounting files were further quality filtered by allowing only identifications with the following criteria: identification in at least two separate samples (not including replicate injections), a process that required at least three high quality unmodified peptides using the Top 3 method, and 2 peptides with at least 4 fragment ions for each protein. All other protein identities were removed. Proteins were first ranked according to coverage across the samples, and then each protein entry was ranked according to the order in which they were run. QC information was added for each sample (batch information, protein concentration, ion counts). First, differences in run-to-run intensity (loading) were adjusted by normalising each run to the sum of top 3 intensities of the proteins up to the point where the sample set reached 10% missing data (we refer to this as 'top-90 normalisation'). ComBat was used to adjust for batch to batch variation [43].

Inforsense software (ID Business Solutions, Guildford, UK) was applied to generated heat maps for the top 150 proteins using both 'top 3 peptide intensity sum' (a proxy for concentration) and peptide concentrations (expressed in fmol) on column calculated from internal standards. Sample-wise correlation plots were created using Inferno RDN (http://omics.pnl.gov/software/infernordn)[44]. Heat maps and correlation plots were inspected for poor samples or injections; those with very low or no ID's and/or poor correlation were removed from the dataset.

Samples were analysed in duplicate and the average intensity values used for analysis. For the purpose of quality control, several analyses were performed. Replicate injections were inspected for consistency in quantitation. To achieve this, an average of the two injections 'top 3 peptide intensity sum' was used and a distance matrix calculated by taking the Euclidian distance between the two injections as a function of the average of the injections. These values were visualised in a heat map, enabling rapid inspection of duplicates with high variance, which likely indicated a technical issue between injections (e.g. sprayer dropout, or failure to inject the correct volume). Data were corrected by applying the following universal rule (Rule 1): for samples with >2-fold between-injection difference in average intensity of all proteins: "report injection one intensity values for all proteins, unless a specific protein was only quantified in injection two,

then include this value for increasing coverage". Injection 1 was selected for consistency as it is not possible to distinguish which run more accurately represents the true abundance.

While the above method was useful in identifying whole samples with poor repeatability between injections, there were cases where the concentrations of individual proteins were highly variable. To assess these cases, a log was created using a custom script, which highlighted those proteins where the ratio between injections was >1.5. Proteins with high frequency of poor measurement stability across all samples were processed according to 'Rule 2: "if the variation between injections is greater than 1.5-fold, take the quantity measured using injection one". The fraction of samples where the ratio between runs had to be >1.5 was 0.5 to apply rule 2. This rule was only applied to 11 proteins in the extended proteome and to 0 proteins in the core proteome.This consistent approach to dealing with large variation in between-run protein measurements was useful in reducing technical variation in the dataset, whilst minimising reductions in proteome coverage. However, we recognise that there remains increased uncertainty in the measurement of proteins treated in this way, and the issue may be minimised in future studies by increased replicate measurements. Mean values were derived from replicate sample injections except for those cases where rule 1 and rule 2 were applied, and those cases where the protein was quantified in only one sample.

Data retrieval and conversion

Data from previous studies of relevant tissues were retrieved from the following sources: the Protein Atlas (www.proteinatlas.org)[45], the HUPO [46, 47] plasma reference set, the reference plasma dataset from the laboratory of Matthias Mann[48], sputum proteomes[2, 21, 24], proteomes of BAL fluid[36, 37, 49-52], exhaled breath condensate[53, 54], pure airway mucus[55], saliva[56-61], macrophage proteomes[62, 63], eosinophil proteomes[64-68], whole neutrophils[69-72], investigations of neutrophil extracellular traps[16, 73, 74], neutrophil microparticles[75], and neutrophil granules[76, 77]. Data were tabulated and identifiers converted to Uniprot format using the Uniprot ID mapping service (http://www.uniprot.org/mapping/) or via DAVID [78, 79]. Redundant, discontinued, merged and incomplete entries (e.g. assignments to protein fragments, pseudogenes, or to non-human proteomes) were either disregarded or were assigned to Uniprot identifiers.

Statistical analyses and informatics

Statistical analyses were performed in R[80], Inferno RDN [44], Microsoft Excel and Minitab [13], using parametric or non-parametric tests as appropriate. Coefficient of variation was calculated for log-normal distributions. Visualisation was performed in Origin 9.1 (http://www.originlab.com/91), R, and Inforsense (IDBS). Venn diagrams were generated using Venny

(http://bioinfogp.cnb.csic.es/tools/venny/)[81]. Tree maps were drawn using Treemap 4.1.2 (http://www.cs.umd.edu/hcil/treemap/)[82]. Pathway analysis, functional enrichment analysis and biological inference were performed using Ingenuity Pathway Analysis software (IPA®, QIAGEN Redwood City). The IPA analyses were performed against a background gene set restricted to Ingenuity Knowledge Base genes from expected sources of sputum proteins, tissues and cell types of the lung or near the lung in homo sapiens. Functional enrichment was calculated via FunRich [83], and the Secreted Protein Database [84] was used to identify secreted proteins. Protein-protein interactions were explored using String (http://string-db.org/)[85]. Ontology was annotated from retrieval via Uniprot[86], and retrieval and enrichment analysis was performed via GOrilla (http://cbl-gorilla.cs.technion.ac.il/)[87] against the homo sapiens proteome and REVIGO (http://revigo.irb.hr/)[88]. The data were clustered by topological data analysis (TDA) [89-93]. TDA provides geometric representation of the relationships between patient data and variables in high-dimensional data sets. TDA structures were generated using the Ayasdi Cure application (Ayasdi, Menlo Park, CA) with a norm correlation metric and two MDS lenses (resolution, 20 bins; gain, ×5.0; equalized).

Power calculations were performed using R-package size[80], assuming two equally sized groups, to achieve 80% power to detect a given fold change (FC), at the FDR-adjusted 5% significance threshold, using a two-sided, two-sample t-test, assuming that the percentage of true null hypotheses, 'p_o', is 95% and 97.5% respectively. Equal variances were assumed for cases and controls. The open source software, variancePartition, was used to identify the drivers of protein measurement variation[94].

RESULTS

Identified proteins

A total of 4,182 proteins were identified in the sputum in ≥1 individual(s): 2,354 proteins in ≥2 individual(s), 284 proteins in ≥40% and 73 in ≥90% of individuals (Supplementary Figure 1). High abundance proteins were generally more frequently identified (Fig. 1), but many high abundance proteins were also identified at lower frequency. These abundant proteins would be expected to be observed across multiple time points in the same patient, although there may be a proportion that are not replicated because of biological variation or where they are near the limits of detection. With consideration for these effects, we have defined the sputum proteome in two ways, the 'core' and 'extended' sputum proteome (supplementary excel file 1). The 284 proteins identified in ≥40% of participants were defined as the 'core sputum proteome' and were used in the statistical analysis. The 'core' proteome represents the most commonly detected proteins within the sputum samples. The cut-off was defined at ≥40%, since at this frequency of identification, the frequency vs. protein rank curve was close to the point

of inflection (Fig. 1), where even a slight increase in the frequency of identification 'cut off', significantly increased the sparsity of the dataset and, hence, the total number of missing values. We also defined an 'extended healthy sputum proteome' dataset consisting of 1666 proteins identified in ≥3 individuals.
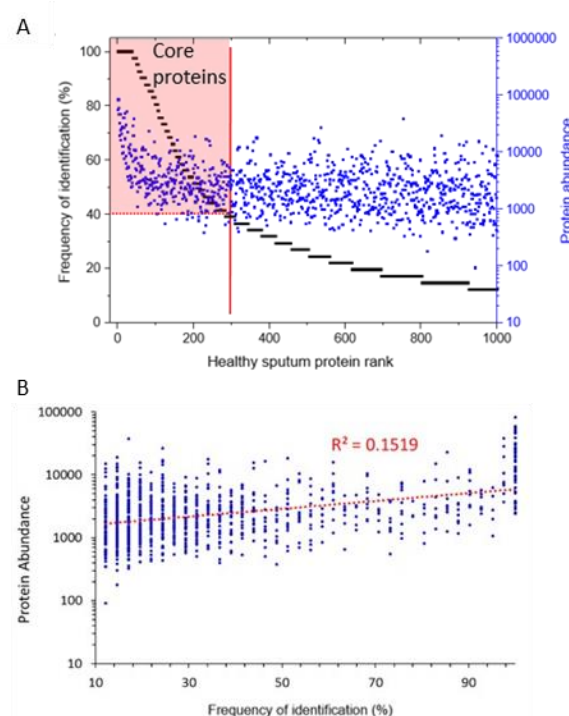

Figure 1. Defining the 'core' sputum proteome from the relationship of abundance of identified proteins and frequency of identification across samples. **A)** illustrates the 40% frequency of detection used as the cut-off for the 'core' sputum proteome. The relationship between rank of frequency of detection and the number of proteins identified (healthy sputum protein rank) is approximately linear between 100% and 40%. This is similar to the relationship between protein abundance and rank of frequency of detection. Red lines indicate 40% cut off points for the 284 protein 'core sputum proteome' and also illustrate that this level is close to the point of inflection of the curve. The 'core sputum proteome' is shaded pink. At this point, increasing the coverage cut off point for analysis significantly increases the sparsity of the dataset and hence the total number of missing values. **B)** The intensity of protein measurement correlated weakly with the frequency of protein identification.

Impact of gender, age, atopy and granulocyte counts

No significant differences in proteomes were observed when comparing age, atopic and non-atopic individuals or males and females (Fig. S2). Furthermore, using an FDR-adjusted t-test, only CLIP-associating protein 1, CLASP1_HUMAN, was found to be significantly different (q=0.04) between males and females. No proteins were significantly different between atopic and non-atopic individuals (q < 0.05). The network shown in Fig. S2 is constructed using multidimensional scaling (MDS) lenses (similarity metric) projected onto a TDA network, representing the structure of the proteomic data. This is an advanced technique for clustering data according to similarity

and was used to explore the shape of the data for impacts of potential covariates. There was a large range of sputum neutrophil counts, and a small number of individuals had counts (>80%) that would be classified as neutrophilia. Compared to the other participants, these individuals had elevated levels of Neutrophil Defensin (Mann Whitney; q=0.02) and borderline results (q=0.06) for neutrophil-associated proteins: leukocyte elastase inhibitor (Serpin B1), MMP9, and S100A8/9, and RHO protein GDP dissociation inhibitor. There was also a weak, but statistically significant, positive correlation between some of the major granule proteins and neutrophil counts as a % of total inflammatory cells (4 of the top 5 proteins with greatest R^2 correlation scores are shown in Fig. 2). The average R^2 for correlation of proteins with neutrophil count was 0.07 and 91 % of proteins had an $R^2 < 0.2$.
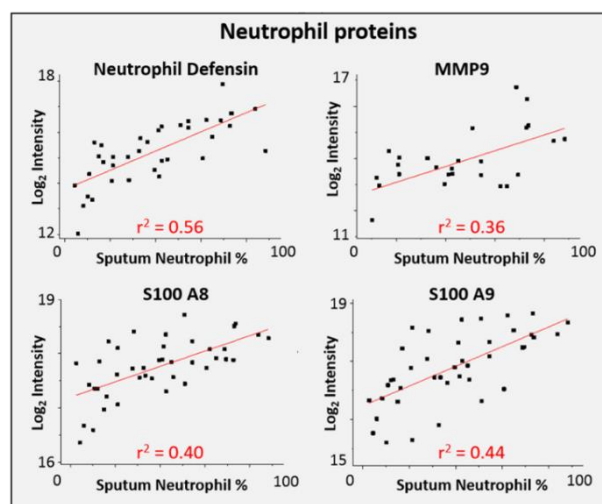
Figure 2. Neutrophil proteins and neutrophil counts. We observed relationships between neutrophil granule proteins and the proportion of neutrophils in sputum. Neutrophil proteins correlated with percentage sputum neutrophil cell counts.

Protein variability, intensity adjustment and measurement accuracy

Protein measurements between samples were visualised in heat-maps before and after intensity adjustment and after batch effect correction (Fig. S3), allowing rapid assessment of fluctuations in instrument performance and systematic variation, e.g. sensitivity and column changes over time and between analytical batches. Effects that can be easily visualised in non-adjusted maps (Fig. S3A) were corrected by normalising the intensity of proteins in each sample to the sum of the measured top3 intensity of the proteins for that sample, up to the level of 10% missing data across all samples; a method which we termed the 'top90' method (Fig. S3B). The top90 adjustment corrected for protein intensity variation more effectively than normalisation to 'total intensity'; avoiding effects of rarely measured protein abundances. ComBat was used to adjust for batch to batch variation (Fig. S3C).

High sample to sample variability was observed in the sputum proteomes. Pooled sputum sample replicates,

which were processed and analysed at regular intervals throughout the acquisition of healthy sputum data, were compared to the participant sputum samples and the matched healthy serum dataset (Fig. 4). Compared to individual sputum samples, pooled samples contained higher numbers of proteins with a lower percentage Coefficient of Variation (CV%). In all datasets, the variability in protein measurements increased as the protein frequency of identification decreased. While the level of variability in sputum pools was low, it was even lower in matched serum samples, thus further indicating that the source of variability arises from the sample type rather than the instrumentation.

We can describe the variability in the sputum as emerging from heterogeneity in the population, assessed by comparing pool variance to population variance, and technical variability; sample processing and measurement in the mass spectrometer. Such technical variability can be assessed through analysis of pooled reference samples and replicate injections, respectively. Furthermore, we used the open source software, variancePartition, to identify the drivers of protein measurement variation. Heterogeneity in the population drives most variance in the measurement of the top 20 proteins with highest coverage across samples, except for IGHA1_human, Immunoglobulin heavy constant alpha 1, whose variance was most attributable to mass spectrometry running batch. Immunoglobulins are highly conserved proteins, difficult to distinguish; minor changes in the running conditions within the mass spectrometer may result in different identity assignment. Across the top 40 proteins with highest coverage across samples, 75% of variability could be attributed to heterogeneity in the population (Fig. 3B).
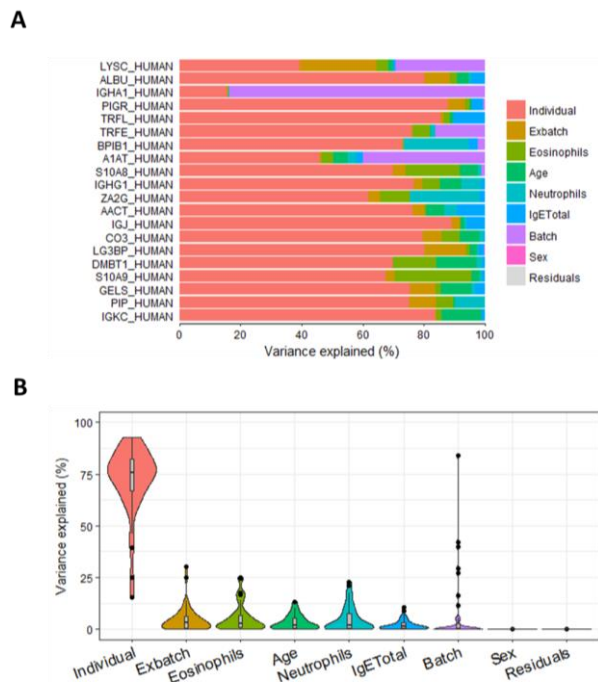
Figure 3. Sources of variability in protein measurement estimated by variancePartition. **A)** the fraction of total variation in measurement of the top 20 proteins by highest coverage in samples, attributable to individual, extraction batch, eosinophil cell count, age, neutrophil cell count, total measured serum IgE, mass spectrometry running batch, sex and residuals. **B)** proteome-wide violin plot of the distribution of variance explained by each variable across the top 40 proteins by highest coverage in samples.

Proteins with high inter-individual population variance are shown in Supplementary Table S1, which illustrates those proteins that had a high CV% in the healthy population but were relatively stable in the pooled samples and injection replicates. Many of the proteins with individual to individual variation are known to have roles in inflammation (S100A proteins[97], A1AT[98]), or are likely the result of salivary contamination of samples (e.g. Amylase[99]).

One of the features used for assessing measurement error was injection repeatability. We defined a poor injection repeat as any protein in a given sample with >1.5-fold difference in measurement between injections. Such variation occurred in ~6% of all quantified sputum proteins, and ~5% of all quantified serum proteins. Variability occurred less frequently in the 'core proteome region' of serum where there was only 1.2% variation across duplicate injections. However, in sputum this value increased to 8.1% (1632 of 20412 individual quantifications) of the identified proteins (~6% in pools). It should be noted that the majority of these poor replicators in sputum occurred where there was a lower frequency of identification (higher rank in Supplementary Tables S2 and S3).
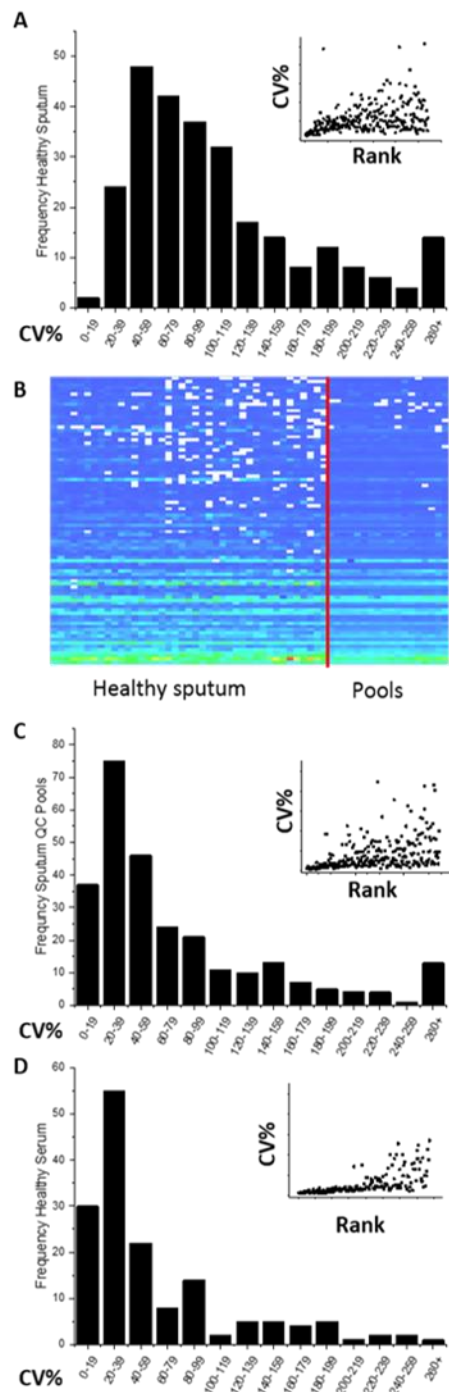
Figure 4. Variability in protein abundance measurements across samples. Frequency histograms represent: on the x axis CV% in increments of bin size 20, and inset scatter plots show CV% vs protein rank (proteins were ranked by order of abundance and frequency of identification across samples). Heatmap **B** illustrates Healthy sputum is highly variable from participant to participant compared to pools. Plots **A** and **C** show variability in protein abundance across healthy sputum samples, and pooled samples respectively; with the pools showing least variability across the different samples. Inset graphs illustrate the variability increase as coverage and abundance decreases. The variability seen in sputum is likely due to sample heterogeneity, and this is contrasted to serum sample measurements in plots **D,** which illustrate the relative homogeneity of that fluid across study participants.
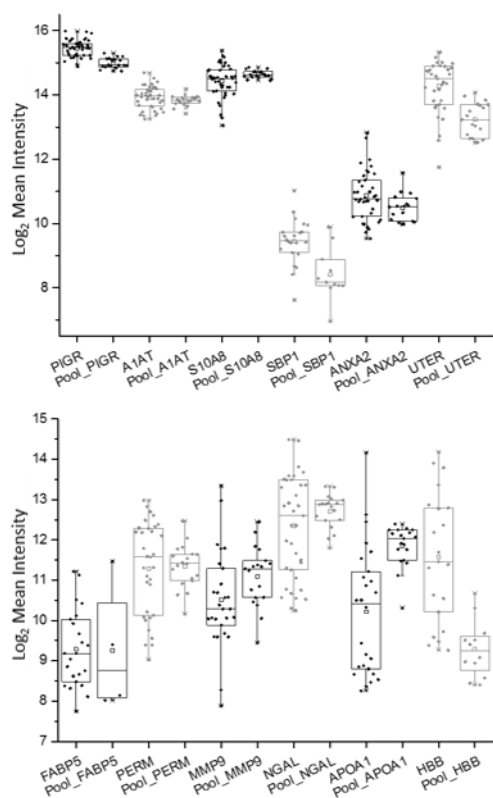
Figure 5. Distribution and variation in measurement of potential inflammatory biomarkers in the healthy population compared to pooled controls. Proteins showed varying levels of spread indicating that the sample size required for statistical power will vary significantly per analyte, with the top graph showing those proteins with lower variability and the bottom, those with higher variability. Note that the pooled samples were taken from asthmatics and non-asthmatics, and as such the means of a number of these proteins will be higher or lower in the pools depending on each protein's role in inflammatory disease. Therefore, pool samples can be used to contrast not only the measurement error but also any potential sub-clinical inflammatory effects in healthy participants.

Proteins that showed variability due to sample handling were identified by their high measurement stability in replicate injections but high variability across the pools (Supplementary Table S2). We observed a number of proteins with poor repeatability of quantification.

Some proteins are difficult to measure with good repeatability by Mass Spectrometry. These poor MS quantifiers showed high CV% in pooled samples and poor replication of quantification across injections. This variability in measurement is likely due to the behaviour of their peptides in HDMS[E] or errors in database searching and quantification (e.g. due to homologues or protein to protein ambiguity) (Supplementary Table S3).

The variability in quantitation was non-uniform throughout the dataset, i.e. it varied on a protein to protein basis. This observation in the samples from healthy participants, which would be used as a control group for comparison with samples from participants with disease, has far reaching implications. This is particularly so in terms of experimental design and statistical power for biomarker discovery using unbiased sampling techniques. In order to explore this phenomenon further, we performed a literature search and generated a database of potential respiratory biomarkers (supplementary excel file 2). The database was used to identify highly cited respiratory proteins associated with disease, and then, in a post-hoc manner and using the data on variability gathered from our experiments, we performed an *in silico* exploratory evaluation of sample size requirements for MS-based biomarker studies for these biomarkers.

Biomarker proteins identified in more than three published studies and also identified in our study were chosen for more detailed analysis. A sub-set of these proteins are presented in Fig. 5. Low variability in the pooled samples, but high variability in the population, indicated that the majority of variability arose from inter-individual variation, while diverse measurements in both the pools and the population indicated that there was a likely influence of experimental variation to consider for that protein.

Twenty-four of the proteins in the biomarker database were observed in ≥20 healthy participants in our study, which allowed a series of sample size calculations to be performed. The calculations quantified the relationship between variability of the samples, given by the standard deviation of the measurements, and the sample size required to achieve 80% power (Table 2 and supplementary Fig. S4). The sample size required for the given statistical power varied between proteins because of a combination of experimental and biological variation. This highlights that while statistical differences for some biomarkers can be reliably identified from sample sizes that are routinely used in proteomic analysis, others require very large sample sizes in order to confidently identify an effect. It is noted that in studies where patient allocation is unbalanced (e.g. 1:2 or 1:3 cases to controls), the total necessary sample size required is greater. A 1:1 allocation provides the most efficient design.

9

Table 2. Sample size per group, to detect 1.5 or 2-fold differences with 80% power.

Protein	Fold Change P0 SD(log2(x))	1.5 95%	97.5%	2 95%	97.5%
PIGR	0.264	9	10	5	6
BPIB1	0.303	11	12	6	6
A1AT	0.349	14	15	7	7
LYSC	0.352	14	15	7	7
ACTB	0.379	16	17	7	8
TRFL	0.478	23	26	10	11
S10A8	0.525	27	30	11	13
S10A9	0.609	36	40	14	16
CO3	0.631	38	43	15	17
SBP1	0.711	48	53	18	20
VTDB	0.714	48	53	18	20
CFAH	0.743	52	58	20	22
ANXA2	0.776	56	63	21	24
PRDX1	0.804	60	67	23	25
UTER	0.825	63	70	24	26
PEDF	0.879	72	80	26	29
B2MG	0.882	72	80	27	30
FABP5	0.956	84	94	31	34
PERM	1.166	124	137	44	49
DEF1	1.168	124	138	44	49
MMP9	1.201	131	146	47	52
NGAL	1.307	155	172	55	61
APOA1	1.588	227	252	80	88
HBB	1.609	233	259	82	91
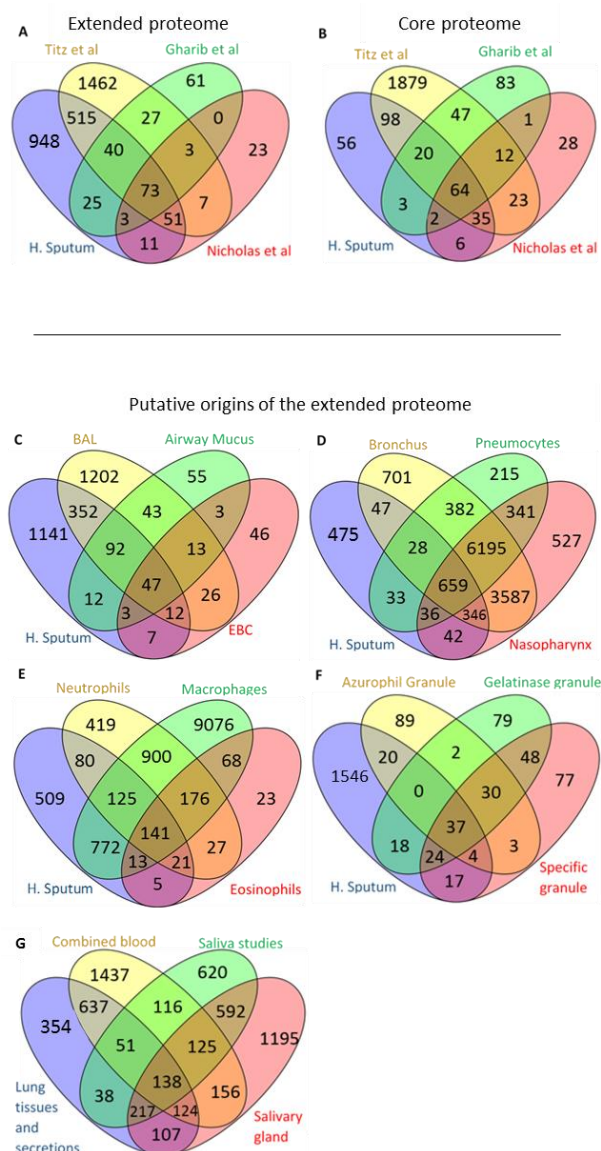
Figure 6. Comparison of coverage of U-BIOPRED sputum with other studies. Including other sputum studies (**A-B**), other proteomic studies of airway fluids and secretions (**C**),

and respiratory tissues found in the protein atlas normal expression database (**D**). U-BIOPRED (Healthy) sputum proteome was compared to known proteomes of granulocytes and macrophages (**E**), and to the proteomes of neutrophil granules (**F**). Comparisons were also made with Saliva and Blood (**G**).

Salivary contamination

We compared the abundance of proteins reported to be salivary proteins against the squamous cell counts from the study participants (Supplementary Figure S5). Although salivary proteins tended to be high when the percentages of squamous cell counts were high, this was not consistent and many participants had low squamous counts but high levels of salivary proteins.

Tissue and cell origins of the sputum proteome

Proteins found in this study have been previously observed in studies of human sputum proteomes. Comparison with studies of induced sputum by Gharib et al.,[24] Nicholas et al.,[2] and Titz [21] (Fig. 6 A to B) showed an extensive overlap of measured proteins. Here, in the extended proteome, we found 63% of the 232 proteins identified by Gharib et al., 81% of the 171 the proteins identified by Nicholas et al., but only 31% of the 2178 proteins found in the study by Titz et al.; reflecting the heterogeneity in sputum proteomes. Overlap of proteins found between two studies particularly with different methodologies increases confidence of the presence of this protein in sputum. Overlap of proteins between more than two studies further supports the identifications. The extended proteome in the current study showed the highest number of proteins shared with other studies, representing an improvement in protein identification, but also added confidence to the proteins identified in few samples by HDMS[E].

The overlap of proteins found between the extended sputum proteome and other tissues and cells indicates potential origins of proteins identified in this study of the sputum proteome. Sputum is a complex biofluid, consisting of proteins of multiple origins and can therefore reflect a complex biological picture. Despite sampling different airway compartments, similar patterns of proteome coverage were also observed when comparing airway mucus[55], exhaled breath condensate (EBC)[53, 54] and studies analysing bronchial alveolar lavage (BAL)[36, 37, 49-52] (Fig. 6C). However, the number of proteins identified here were only of similar magnitude to those in the BAL study. This may be partially due to improvements in sensitivity of protein measurement. Approximately 20% of proteins identified in the BAL study were also identified in the current study, which likely reflects an overlap in the high confidence detection of high abundance proteins, and a variability between studies in sampling of less abundant proteins. Results also showed extensive overlap with proteins identified from respiratory tissue analyses in the Protein Atlas[45] (Fig. 6D). There was also overlap in the identifications of measured proteins to those of proteomes measured for eosinophils, macrophages and granulocyte (Fig. 6E) and neutrophil

granules (Fig. 6F). Identification of proteins from all major tissues and cell types of the airways highlights not only the complexity of sputum as a clinical fluid, but also its utility for accessing lung biology. Such results were also reflected in our core sputum proteome, where extensive overlap was observed across a variety of tissues and/or biofluids (Supplementary Fig. S6). The interpretation of these results is limited due to the differences in the sensitivity and variability of the protein measurement techniques. In most cases, the proteins identified in the studies are not tissue or cell specific, for example, the proteome of the compared macrophage data set covers half of the genome, which suggests that most of these proteins are not macrophage specific. With increasing sensitivity in protein measurement techniques, there is a corresponding increase in the proportion of low abundance proteins, therefore, overlap of proteins detected in studies reporting fewer proteins may be more useful in finding the biological origins of the proteins. In studies where fewer proteins are reported, these likely reflect high abundance, easy to detect proteins and do not include low abundance proteins which are harder to detect.
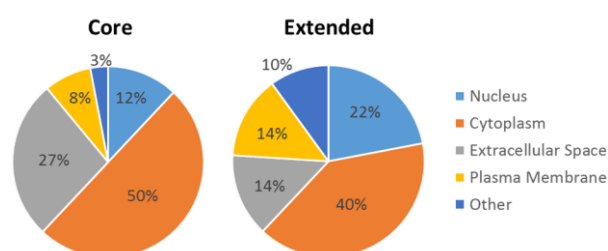
Figure 7 Subcellular localisation prediction results from IPA analyses of the proteins from the core and extended sputum proteome against a background gene set restricted to Ingenuity Knowledge Base genes from expected sources of sputum proteins.

To further understand the protein composition of sputum, we also investigated possible systemic or non-respiratory origins of our identified proteins through comparison to other relevant biofluids. We identified ~ 50% overlap with proteomes of sputum, the upper airway, saliva and blood, similar to that reported previously[2, 61] (Fig. 6G).

Functional analysis

Subcellular localisation of proteins was analysed using Ingenuity Pathways Analysis (IPA) (Fig. 7). 27% of the proteins in the core sputum proteome were predicted to be extracellular or secreted, with a further 8% predicted to be integral membrane or cell surface proteins. 50% of identities were predicted to be cytoplasmic proteins.

The LXR/RXR pathway refers to liver X receptor (LXR) which is activated by oxysterol ligands to bind retinoid X receptors (RXRs). Resultant LXR-RXR heterodimers bind LXR response elements and regulate expression of genes involved in inflammation, metabolism and cholesterol metabolism. The FXR/RXR pathway refers to the bile-acid concentration-mediated farnesoid X receptor (FXR) and RXR regulation of lipid metabolism. Integrin linked kinase (ILK) signalling refers to the ILK-mediated control of cytoskeleton remodelling. eNOS Signalling refers to the mechanism of Nitric Oxide (NO) production by the endothelial NO Synthase (eNOS).

Enrichment and pathway analyses for the core sputum proteome and extended sputum proteome were performed using Funrich, Go-Rilla and IPA. As gene ontology mapping presents a high-level of redundant terms, we utilised REVIGO to collapse and summarise like terms as treemaps. The results of these GO enrichment analyses for the core sputum proteome are shown in Supplementary Fig. S7, A for cellular components, B for molecular function, and C for biological processes. Analyses of the extended sputum proteome were also performed and shown in Fig. S7 D-E. The size of each individual square represented in each treemap is proportional to the $-\log_{10}$ p value for the enrichment of that category, measured by Fisher's exact test. The ontologies are grouped by related terms and defined by colour, providing a landscape overview of induced sputum. Enrichment was observed for vesicle related components; GORilla q=1.92E^{-7}, with 111 proteins identified with the GO term 'vesicle', however, proteins have multiple GO terms and these results only suggest an enrichment of proteins originating from vesicles. Many granulocyte functions involve vesicle formation, such as the release of extracellular vesicles, some with antibacterial effect, released during spontaneous death of neutrophils[131]. Enrichment was also observed for extracellular proteins, (immune) receptor and antigen binding functions and dominated by processes involved in homeostasis, and mucosal and innate immunity. Further analysis of enriched biological pathways showed that the top IPA canonical pathways are thematically similar (Table 3), with cell migration and tissue organisation (integrin linked kinase signalling, actin cytoskeleton signalling, leukocyte extravasation signalling), innate immunity (acute phase response, complement) and regulation of cytoskeleton, extracellular matrix (ECM) remodelling and inflammation (e.g. FXR/RXR (farnesoid X receptor / retinoid X receptor) or LXR (liver X receptor)/RXR, RhoA - Ras homolog A, RhoGDI - Rho GDP-dissociation inhibitor) being the most enriched pathways and, to a lesser extent, those involved in energy metabolism (e.g. glycolysis). Similar trends were observed when exploring the association of proteins with specific functions and diseases, with inflammatory responses and immune cell migration dominating the functional enrichment categories (Table 4). These enriched pathways and functions were also conserved in the extended sputum proteome dataset (Supplementary Tables S4 and S5).

Investigating protein interaction networks (http://string-db.org) in our core sputum proteome dataset (Supplementary Fig. S8), we observed a number of small groups of interacting protein partners and significantly one large interaction network, highlighting the large number of functional

11

relationships that are experimentally accessible in the core sputum proteome.

Table 3. Top 20 Enriched canonical pathways in the core sputum proteome set.

Ingenuity Canonical Pathways	Enrichment P-value	Total number matched proteins
LXR/RXR Activation	$3.16E^{-12}$	17
Glycolysis I	$2.00E^{-11}$	9
Acute Phase Response Signalling	$6.31E^{-11}$	17
FXR/RXR Activation	$5.01E^{-10}$	15
Gluconeogenesis I	$7.76E^{-10}$	8
Actin Cytoskeleton Signalling	$2.69E^{-07}$	15
ILK Signalling	$2.00E^{-06}$	13
Primary Immunodeficiency	$3.89E^{-06}$	7
Pyruvate Fermentation to Lactate	$1.45E^{-05}$	3
Complement System	$1.70E^{-05}$	6
RhoGDI Signalling	$2.95E^{-05}$	11
Hematopoiesis from Pluripotent	$3.16E^{-05}$	6
Leukocyte Extravasation Signalling	$3.89E^{-05}$	12
RhoA Signalling	$3.89E^{-05}$	9
Clathrin-mediated Endocytosis	$5.50E^{-05}$	11
Glucocorticoid Receptor Signalling	$7.08E^{-05}$	14
Cell Cycle: G2/M DNA Damage	$8.13E^{-05}$	6
Epithelial Adherens Junction	$1.69E^{-04}$	9
Coagulation System	$1.90E^{-04}$	5
eNOS Signalling	$1.99E^{-04}$	9

Table 4 Top 20 enriched functions and diseases found by IPA of the core sputum proteome.

Disease and function category hierarchy	Specific annotation	Enrichment p value	Number of matched proteins
Infectious Diseases, Inflammatory Disease, Respiratory Disease	Severe acute respiratory syndrome	$1.92E^{-09}$	14
Infectious Diseases	Viral Infection	$4.17E^{-09}$	18
Inflammatory Response	Inflammatory response	$4.40E^{-08}$	18
Connective Tissue Disorders, Inflammatory Disease, Skeletal and Muscular Disorders	Rheumatic Disease	$9.30E^{-08}$	19
Connective Tissue Disorders, Inflammatory Disease, Skeletal and Muscular Disorders	Arthritis	$1.30E^{-07}$	18
Cellular Movement, Immune Cell Trafficking	Leukocyte migration	$2.22E^{-07}$	20
Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking, Inflammatory Response	Cell movement of phagocytes	$2.87E^{-07}$	16
Immunological Disease	Systemic autoimmune syndrome	$3.05E^{-07}$	16
Inflammatory Disease	Chronic inflammatory disorder	$5.92E^{-07}$	16
Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking	Cell movement of leukocytes	$6.79E^{-07}$	19
Connective Tissue Disorders, Immunological Disease, Inflammatory Disease, Skeletal and Muscular Disorders	Rheumatoid arthritis	$1.02E^{-06}$	15
Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking	Cell movement of myeloid cells	$1.73E^{-06}$	15
Cell-To-Cell Signalling and Interaction, Haematological System Development and Function, Immune Cell Trafficking	Adhesion of immune cells	$3.58E^{-06}$	13
Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking, Inflammatory Response	Chemotaxis of phagocytes	$5.26E^{-06}$	12
Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking, Inflammatory Response	Chemotaxis of leukocytes	$6.22E^{-06}$	14
Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking, Inflammatory Response	Chemotaxis of myeloid cells	$8.82E^{-06}$	12
Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking	Cell movement of granulocytes	$1.06E^{-05}$	11
Cellular Assembly and Organization	Formation of rosettes	$1.45E^{-05}$	3
Organismal Survival	Organismal death	$1.86E^{-05}$	7
Cellular Movement, Haematological System Development and Function, Immune Cell Trafficking, Inflammatory Response	Cell movement of neutrophils	$2.08E^{-05}$	9

DISCUSSION

This study provides the first large-scale analysis of sputum using MS^E, a data-independent proteomics approach. To our knowledge, it provides the most comprehensive description to date of airway lining fluid proteins and, thus, significantly advances on the sputum proteome we have previously reported[2]. The study has identified 284 proteins in the core healthy sputum proteome that are reliably and repeatedly measured and 1666 proteins in the extended proteome, additionally detailing less repeatedly measured proteins that are usually also less abundant. We have assessed the biological relevance of the proteome, particularly in the context of enrichment against the wider human proteome and have highlighted that secreted proteins and vesicular components are highly enriched in this biofluid.

Proteomic studies of sputum to date have been small but, nevertheless, useful. Our previous study[2] identified 191 proteins in the sputum of a single individual using 2-dimensional gel electrophoresis and mass spectrometry which is biased towards high abundance proteins and not amenable to high throughput analysis. Comparison of our results with sputum studies by other authors shows excellent overlap in detected proteins, but also a number of proteins unique to each study. Gharib *et al.,* applied shotgun mass spectrometry to assess 5 healthy and 10 asthmatic participants [24 25], and identified 254 proteins in all participants' sputum. Using a non-parametric test developed by the authors, called the spectral index[26], they found 17 proteins whose concentrations were significantly different between asthma and health, including serpin peptidase inhibitor (SRPINA1) and secretoglobin (SCGB1A1, also known as Clara cell 10-kD protein) that were increased and decreased in concentration, respectively in asthma. Titz *et al.,*[21] conducted an impressive study applying LC–MS/MS of tryptic peptides labelled with Tandem Mass Tags (TMT) to sputum samples of 216 participants equally composed of healthy non-smokers, healthy never smokers, COPD patients and current smokers. Relative quantification was achieved by sequentially

measuring the proteome of demographically matched samples (1 from each cohort per run) against a pooled reference sample. They analysed differential protein expression only in proteins detected in at least 2/3 of samples per study group and reported proteins differentially abundant between groups but did not report the total proteomes. This approach pointed to 15 proteins differentially abundant when comparing patients with COPD and current smokers and many more when compared with non-smokers.

Differences in reported proteins between studies can be explained by differences between individual phenotypes; biological variability. Differences are also attributable to differences in sample preparation and analysis methods. Likewise, where there are similarities in techniques used, there are large overlaps in proteins reported. Highlighting this paradigm, mucins 5A and 5C and the IgGFc-binding protein, which are all major components of airway mucus [132, 133,] were not identified in our study, likely due to due to post-digestion filtration of samples to remove large indigestible substances e.g. mucopolysaccharides, thereby removing these proteins from the analysis.

Protein detection and quality control

The study design and attention to quality control has allowed us to investigate numerous experimental and measurement effects in our dataset: technical reproducibility in measurement of each protein, inter-individual variability of each protein and identification of salivary contamination of sputum samples. The semi-stochastic nature of peptide sampling in MS-based proteomic approaches, even when using data independent methods, often results in highly abundant proteins being measured more reproducibly, with proteins of lower abundance being identified less frequently, leading to increased sparsity within a dataset. This is most evident in larger studies, where the proportion of proteins identified across all samples is lower than in smaller studies. These effects were visible in both the serum and sputum datasets in the current study, and as such, influenced our approach to developing our 'Top-90 intensity normalisation' strategy. In a data matrix ranked by protein intensity and frequency of identification, the low frequency region was seen to be more variable between samples, both in the number of protein identities per sample and their intensity measurement. Such variation can influence normalisation strategies based upon total MS intensity. We have found that using the region with the greatest coverage for adjusting samples, the top90 method, is an effective method of normalisation for protein load and intensity.

Comparison of variability across injection replicates and pooled QC samples within the healthy sputum proteome dataset allowed the precision of measurement for specific proteins to be assessed. Notably, measurements of some proteins were reproducible across samples, whereas other proteins showed poor replicability across sample replicates and pooled QC samples, indicating that these proteins are intrinsically recalcitrant to precise measurement by HDMS[E]. As expected, there was a trend for proteins with lower intensity to show higher variability since they are generally represented by fewer peptides and are more prone to interference due to noise. In this study we chose to define a core proteome using proteins detected in ≥40 % of samples representing proteins reproducibly measured. However, we recognise that the proteome presented here will require verification in future studies.

By systematically examining the variability in pools, samples and replicate injections, we were able to identify potential sources of variation as either: sample heterogeneity, sample processing effects, and MS measurement or post processing effects. Proteins that display high CV of measurement across pools, but good replication between injections, are likely to be poor quantifiers due to sample preparation and are likely to be unstable or variably modified. The majority of these are at the lower end of the coverage spectrum. Many of the variable proteins were also identified as membrane or centriolar proteins, which often contain hydrophobic protein regions that can affect protein digestion or peptide extraction efficiencies, thereby influencing the MS measurement. In a similar fashion, those proteins that were stable in the pools, but were variable in injection replicates, are likely to be poor MS quantifiers or liable to post-processing errors. Many of these have potential homologs which could interfere with their *in silico* quantitation. The results of our QC analysis suggest that variation of proteins in pooled QC samples and sample replicate measurements need to be routinely assessed across sample populations for individual proteins and for different biofluids. This will be critical in developing targeted biomarker assays and designing large discovery projects.

Protein measurement and concentration variability and effect on statistical power

Determination of normal/healthy levels of proteins is essential for the identification of potential disease biomarkers. It is, therefore, crucial to understand both the technical and biological variation affecting measurement of this baseline proteome. In the current study, the measurements of a number of 'core sputum proteome' proteins were highly stable in repeat analyses of pooled sputum samples, pointing to the robustness of the MS analysis, but inter-individual variability was significant (Supplementary table S1), in stark contrast to the low inter-individual variability in serum from the same individuals.

Based upon respiratory protein biomarkers previously identified from other studies and the variability of the proteins observed in our healthy population study, we calculated the sample size required to observe a specific effect against that baseline in the context of a disease setting. By parallel reference to our pooled sample measurements, we inferred how much of the variance can be attributed to processing and/or technical

measurement and how much is a consequence of inter-individual heterogeneity. The greater the variability in the measurement, the greater the sample size required to achieve 80% power (Table 2). For example, quantitation of modest (1.5-fold) differences for a highly variable protein, such as haemoglobin B (HBB), requires >500 patients. By contrast, differential expression of proteins such as polymeric immunoglobulin receptor may be quantifiable with as few as 20 patients. In the case of HBB, high sample-to-sample variation was observed, although variation of HBB across the pooled samples was low. As highlighted above, saliva is often a contaminant in healthy individuals who have difficulty producing an adequate sputum sample. Since saliva is susceptible to contamination with blood as a result of gum disease [129, 130], HBB may be a consequence of oral-derived blood contaminating the sputum. The other highly variable proteins, neutrophil gelatinase associated lipocalin, MMP9, and neutrophil defensin are all known to be involved in inflammation, and several of these seen in the current study correlated with sputum neutrophil numbers.

Protein biomarkers have been reported in previous respiratory studies, however, high variability in protein concentrations, in combination with small sample size[30] can result in false positive biomarker identifications. Additionally, as we have shown, the protein to protein differences in variability means that predicting sample size to sufficiently power biomarker investigations is difficult to achieve without extensive prior investigation. These problems are likely strong contributors to the poor record of translating many biomarkers to the clinic[27, 28].

Confounding factors in sputum analysis

Previous studies have shown intermediate levels of inflammation and even tissue remodelling in healthy atopic individuals relative to those with a clear diagnosis of asthma (Djukanovic ERJ). However, in the current study, there were no significant differences between atopic and non-atopic participants, possibly because their allergic status was too mild or the allergic processes in the airways were not active at the time of sampling.

A small number of individuals displayed neutrophilia (>80% neutrophils) which was reflected by observed differences in neutrophil proteins measured in these study participants. Neutrophilic inflammation is often a hallmark of infection, disease exacerbations, or severity of chronic airway diseases like asthma and COPD [124, 125], but is often confounded by smoking status and steroid treatment, influencing neutrophil half-life, activity and migration [126-128]. A number of neutrophil proteins correlated with neutrophil cell counts (Fig. 2); neutrophil defensin, S100A8, S100A9 and MMP9. Concentrations of the major neutrophil granule protein, myeloperoxidase only weakly correlated with sputum % neutrophil cell counts. Although neutrophil cell counts correlate specifically with these neutrophil derived proteins, we do not observe a large effect across the rest of the core or extended proteome as shown in Figure 3. As such, these proteins are useful biomarkers of neutrophilia whilst not compromising the potential identification of other disease signatures.

Salivary contamination is an inevitable confounding factor in sputum analyses, particularly in healthy individuals and other participants where sputum induction is less successful. We compared squamous cell counts with the expression of proteins found in saliva in previous studies and found mixed correlations. Squamous cell count does not perfectly reflect salivary contamination of sputum samples, however, the salivary proteins studied in these comparisons may additionally originate from other tissues, not just saliva. Further investigation is required to determine better markers of salivary contamination in sputum.

Origins and biological context of the sputum proteome

We investigated the predicted subcellular localisation of the proteins from the core and extended sputum proteomes (Fig. 7 & S7). A large portion of the proteins measured in the sputum proteome were reported to be cytoplasmic by GOrilla analysis. Since our analysis is of the supernatant from induced sputum, where cells and cell debris are removed prior to analysis, this was expected. In addition, the supernatant would also include granulocytes, which contain granules of cytoplasmic proteins for release into the sputum.

When comparing the sputum proteome in the current study with those from studies of airway tissue and fluid samples, significant numbers of identified proteins were seen to overlap. The number of proteins for which a potential cell or tissue origin could not be assigned was low. Significant numbers of the identified proteins likely have origins in other tissues of the body: for example, acute phase proteins are produced largely in the liver and enter the lungs via capillaries together with other plasma proteins. Functional analysis of the healthy sputum proteome identified proteins associated with innate immune defence, inflammatory responses via complement and acute phase proteins, phagocytic cells (macrophages and neutrophils) and, to a lesser extent, eosinophils.

A number of signalling pathways were enriched in sputum, including LXR, a member of the nuclear receptor family of transcription factors that are closely related to nuclear receptors, such as the peroxisome proliferator-activated receptors (PPARs), FXR, also known as bile acid receptor, RXR, that is activated by retinoic acid. FXR/RXR are known to be important in macrophage lipid metabolism and inflammation[133, 135, 136]. Such regulatory pathways have complex roles in inflammatory biology: FXR inhibits inflammation[137], while LXR agonists have been shown to increase airway reactivity and smooth muscle growth in an asthma model[138]. LXR has also been implicated in counter-regulation of toll like receptor induced inflammatory

responses[139, 140], which are involved in inflammatory respiratory diseases[146]. We also observed an enrichment of the RhoA and RhoGDI signalling pathways, known to be involved in hyper-responsiveness in asthma[143, 144], and superoxide generation in macrophages[145], respectively.

An interesting observation was the significant enrichment of vesicles and vesicular components in the sputum (Fig. S7). These small membraned particles, either described as exosomes (nanovesicles) or ectosomes (microparticles), are released by multiple cell types, including immune cells, and have been reported in sputum and lung secretions[146]. Secretory vesicles have recently become an area of interest for their potential pro and anti-inflammatory functions[146]. In response to non-specific complement mediated inflammation neutrophils produce ectosomes that are coated in and loaded with proteins often associated with granules[148]. In addition to neutrophil derived vesicles, antigen-loaded exosomes from mast cells, dendritic cells, epithelial cells and T lymphocytes have been highlighted as being potentially important for allergy[149], and eosinophils produce cytokine containing vesicles e.g. degranulation, that may also be important for asthma (reviewed in Spencer *et al*150).

Secretory vesicles contain high levels of cytoplasmic proteins[77]. For example, exosomes isolated in the BAL of asthmatics are enriched for inflammatory leukotriene production[149] and may help explain the large numbers of cytosolic proteins we have identified in sputum. In addition to the protein loading capacity of these vesicles, functions involving the transport of nucleic acids have also been identified. For example, microRNA-loaded vesicles have been highlighted as potentially important in asthma and inflammatory lung disease signalling[151], and have been shown to be different between BAL samples of healthy and asthmatic individuals[152].

CONCLUSIONS

The mapping of the healthy human sputum proteome in the current study has considered experimental and technical variability, population variance driven by daily exposure of the lung to the external environment, and sample complexity due to multiple potential protein origins, effects of cellular composition and potential for contribution of vesicular components. Functionally, homeostasis and defence mechanisms dominate the measured sputum proteome. The specific experimental and technical variability of the applied methodology must not be underestimated and there are implications for minimum sample sizes required for determining differential abundance between groups with statistical power. The comprehensive approach we have used for this analysis of the healthy sputum proteome provides a good comparator dataset for proteomes from patients suffering from inflammatory lung disease.

Edwards, Asthma UK, London, UK; Rosalia Emma, Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy; Magnus Ericsson, Karolinska University Hospital, Stockholm, Sweden; Breda Flood, European Federation of Allergy and Airways Diseases Patient's Associations, Brussels, Belgium; Hector Gallart, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; Cristina Gomez, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; Kerry Gove, NIHR Southampton Respiratory Biomedical Research Unit and Clinical and Experimental Sciences, Southampton, UK; Neil Gozzard, UCB, Slough, UK; John Haughney, International Primary Care Respiratory Group, Aberdeen, Scotland; Lorraine Hewitt, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK; Jens Hohlfeld, Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany; Cecile Holweg, Respiratory and Allergy Diseases, Genentech, San Francisco, USA; Richard Hu, Amgen Inc. Thousand Oaks, USA; Sile Hu, National Heart and Lung Institute, Imperial College, London, UK; Juliette Kamphuis, Longfonds, Amersfoort, The Netherlands; Erika J. Kennington, Asthma UK, London, UK; Dyson Kerry, CromSource, Stirling, UK; Hugo Knobel, Philips Research Laboratories, Eindhoven, The Netherlands; Johan Kolmert, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; Maxim Kots, Chiesi Pharmaceuticals, SPA, Parma, Italy; Scott Kuo, National Heart and Lung Institute, Imperial College, London, UK; Maciej Kupczyk, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; Bart Lambrecht, University of Gent, Gent, Belgium; Saeeda Lone-Latif, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; Matthew J. Loza, Janssen R&D, USA; Lisa Marouzet, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK; Jane Martin, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK; Sarah Masefield, European Lung Foundation, Sheffield, UK; Caroline Mathon, Centre of Allergy Research, Karolinska Institutet, Stockholm, Sweden; Sally Meah, National Heart and Lung Institute, Imperial College, London, UK; Andrea Meiser, Data Science Institute, Imperial College, London, UK; Leanne Metcalf, previously at: Asthma UK, London, UK; Maria Mikus, Science for Life Laboratory and The Royal Institute of Technology, Stockholm, Sweden; Montse Miralpeix, Almirall, Barcelona, Spain; Philip Monk, Synairgen Research Ltd, Southampton, UK; Shama Naz, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; Ben Nicholas, University of Southampton, Southampton, UK; Peter Nilsson, Science for Life Laboratory and The Royal Institute of Technology, Stockholm, Sweden; Jörgen Östling, AstraZeneca, Mölndal, Sweden; Antonio Pacino, Lega Italiano Anti Fumo, Catania, Italy; Susanna Palkonen, European Federation of Allergy and Airways Diseases Patient's Associations, Brussels, Belgium; Stelios Pavlidis, National Heart and Lung Institute, Imperial College, London, UK; Giorgio Pennazza, University of Rome 'Tor Vergata', Rome Italy; Anne Petrén, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; Sandy Pink, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK; Anthony Postle, University of Southampton, UK; Pippa Powell, European Lung Foundation, Sheffield, UK; Malayka Rahman-Amin, Previously at: Asthma UK, London, UK; Navin Rao, Janssen R&D, USA; Lara Ravanetti, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; Emma Ray, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK; Stacey Reinke, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; Leanne Reynolds, previously at: Asthma UK, London, UK; John Riley, Respiratory Therapeutic Unit, GSK, London, UK; Martine Robberechts, MSD, Brussels, Belgium; Amanda Roberts, Asthma UK, London, UK; Kirsty Russell, National Heart and Lung Institute, Imperial College, London, UK; Michael Rutgers, Longfonds, Amersfoort, The Netherlands; Marco Santoninco, University of Rome 'Tor Vergata', Rome Italy; Corinna Schoelch, Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany; James P.R. Schofield, Centre for Proteomic Research, Institute for Life Sciences, University of Southampton, Southampton, UK; Marcus Sjödin, Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; Paul J. Skipp, Centre for Proteomic Research, Institute for Life Sciences, University of Southampton, Southampton, UK; Barbara Smids, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; Caroline Smith, NIHR Southampton Respiratory Biomedical Research Unit, Southampton, UK; Jessica Smith, Asthma UK, London, UK; Katherine M. Smith, University of Nottingham, UK; Doroteya Staykova, University of Southampton, Southampton, UK; Kai Sun, Data Science Institute, Imperial College, London, UK; John-Olof Thörngren, Karolinska University Hospital, Stockholm, Sweden; Bob Thornton, MSD, USA; Jonathan Thorsen, COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark; Marianne van de Pol, Academic Medical Centre, University of Amsterdam, Amsterdam ,The Netherlands; Marleen van Geest, AstraZeneca, Mölndal,

Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005949 and 10.6019/PXD005949.

REFERENCES

(1) Nicholas B, Djukanovic R. Induced sputum: a window to lung pathology. Biochem Soc Trans 2009; 37:868-72.

2) Nicholas B, Skipp P, Mould R, Rennard S, Davies DE, O'Connor CD, et al. Shotgun proteomic analysis of human-induced sputum. Proteomics 2006; 6:4390-401.

3) Nicholas BL, O'Connor CD, Djukanovic R. From proteomics to prescription-the search for COPD biomarkers. Copd 2009; 6:298-303.

4) Hastie AT, Moore WC, Li H, Rector BM, Ortega VE, Pascual RM, et al. Biomarker surrogates do not accurately predict sputum eosinophil and neutrophil percentages in asthmatic subjects. J Allergy Clin Immunol 2013; 132:72-80 e12.

5) Westerhof GA, Korevaar DA, Amelink M, de Nijs SB, de Groot JC, Wang J, et al. Biomarkers to identify sputum eosinophilia in different adult asthma phenotypes. Eur Respir J 2015; 46:688-96.

6) Tak T, Hilvering B, Tesselaar K, Koenderman L. Similar activation state of neutrophils in sputum of asthma patients irrespective of sputum eosinophilia. Clin Exp Immunol 2015; 182:204-12.

7) Bergquist M, Jonasson S, Hjoberg J, Hedenstierna G, Hanrieder J. Comprehensive multiplexed protein quantitation delineates eosinophilic and neutrophilic experimental asthma. BMC Pulm Med 2014; 14:110.

8) Chua JC, Douglass JA, Gillman A, O'Hehir RE, Meeusen EN. Galectin-10, a potential biomarker of eosinophilic airway inflammation. PLoS One 2012; 7:e42549.

9) Emmanouil P, Loukides S, Kostikas K, Papatheodorou G, Papaporfyriou A, Hillas G, et al. Sputum and BAL Clara cell secretory protein and surfactant protein D levels in asthma. Allergy 2015; 70:711-4.

10) Yilmaz I, Bayraktar N, Ceyhan K, Secil D, Yuksel S, Misirligil Z, et al. Evaluation of vascular endothelial growth factor-A and Endostatin levels in induced sputum and relationship to bronchial hyperreactivity in patients with persistent allergic rhinitis monosensitized to house dust. Rev Port Pneumol (2006) 2015.

11) Desai D, Gupta S, Siddiqui S, Singapuri A, Monteiro W, Entwisle J, et al. Sputum mediator profiling and relationship to airway wall geometry imaging in severe asthma. Respir Res 2013; 14:17.

12) Mastalerz L, Celejewska-Wojcik N, Wojcik K, Gielicz A, Cmiel A, Ignacak M, et al. Induced sputum supernatant bioactive lipid mediators can identify subtypes of asthma. Clin Exp Allergy 2015; 45:1779-89.

13) Yan X, Chu JH, Gomez J, Koenigs M, Holm C, He X, et al. Noninvasive analysis of the sputum transcriptome discriminates clinical phenotypes of asthma. Am J Respir Crit Care Med 2015; 191:1116-25.

14) Gao J, Ohlmeier S, Nieminen P, Toljamo T, Tiitinen S, Kanerva T, et al. Elevated sputum BPIFB1 levels in smokers with chronic obstructive pulmonary disease: a longitudinal study. Am J Physiol Lung Cell Mol Physiol 2015; 309:L17-26.

15) Tangedal S, Aanerud M, Persson LJ, Brokstad KA, Bakke PS, Eagan TM. Comparison of inflammatory markers in induced and spontaneous sputum in a cohort of COPD patients. Respir Res 2014; 15:138.

16) Grabcanovic-Musija F, Obermayer A, Stoiber W, Krautgartner WD, Steinbacher P, Winterberg N, et al. Neutrophil extracellular trap (NET) formation characterises stable and exacerbated COPD and correlates with airflow limitation. Respir Res 2015; 16:59.

17) Zuiker RG, Kamerling IM, Morelli N, Calderon C, Boot JD, de Kam M, et al. Reproducibility of biomarkers in induced sputum and in serum from chronic smokers. Pulm Pharmacol Ther 2015; 33:81-6.

18) Casado B, Iadarola P, Pannell LK, Luisetti M, Corsico A, Ansaldo E, et al. Protein expression in sputum of smokers and chronic obstructive pulmonary disease patients: a pilot study by CapLC-ESI-Q-TOF. J Proteome Res 2007; 6:4615-23.

19)   Nicholas BL, Skipp P, Barton S, Singh D, Bagmane D, Mould R, et al. Identification of lipocalin and apolipoprotein A1 as biomarkers of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2010; 181:1049-60.

20)   Baraniuk JN, Casado B, Pannell LK, McGarvey PB, Boschetto P, Luisetti M, et al. Protein networks in induced sputum from smokers and COPD patients. Int J Chron Obstruct Pulmon Dis 2015; 10:1957-75.

21)   Titz B, Sewer A, Schneider T, Elamin A, Martin F, Dijon S, et al. Alterations in the sputum proteome and transcriptome in smokers and early-stage COPD subjects. J Proteomics 2015; 128:306-20.

22)   Suojalehto H, Kinaret P, Kilpelainen M, Toskala E, Ahonen N, Wolff H, et al. Level of Fatty Acid Binding Protein 5 (FABP5) Is Increased in Sputum of Allergic Asthmatics and Links to Airway Remodeling and Inflammation. PLoS One 2015; 10:e0127003.

23)   Terracciano R, Preiano M, Palladino GP, Carpagnano GE, Barbaro MP, Pelaia G, et al. Peptidome profiling of induced sputum by mesoporous silica beads and MALDI-TOF MS for non-invasive biomarker discovery of chronic inflammatory lung diseases. Proteomics 2011; 11:3402-14.

24)   Gharib SA, Nguyen EV, Lai Y, Plampin JD, Goodlett DR, Hallstrand TS. Induced sputum proteome in healthy subjects and asthmatic patients. J Allergy Clin Immunol 2011; 128:1176-84 e6.

25)   Diamandis EP. The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? BMC Med 2012; 10:87.

26)   Fu X, Gharib SA, Green PS, Aitken ML, Frazer DA, Park DR, et al. Spectral index for assessment of differential protein expression in shotgun proteomics. J Proteome Res. 2008;7:845–54

27)   Drucker E, Krapfenbauer K. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. Epma j 2013; 4:7.

28)   Frantzi M, Bhat A, Latosinska A. Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. Clin Transl Med 2014; 3:7.

29)   Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. Cancer Res 2012; 72:6097-101.

30)   Mischak H, Allmaier G, Apweiler R, Attwood T, Baumann M, Benigni A, et al. Recommendations for biomarker identification and qualification in clinical proteomics. Sci Transl Med 2010; 2:46ps2.

31)   Skates SJ, Gillette MA, LaBaer J, Carr SA, Anderson L, Liebler DC, et al. Statistical Design for Biospecimen Cohort Size in Proteomics-based Biomarker Discovery and Verification Studies. Journal of Proteome Research 2013; 12:5383-94.

32)   Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, et al. Quantitative proteomic analysis by accurate mass retention time pairs. Anal Chem 2005; 77:2187-200.

33)   Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics 2006; 5:144-56.

34)   Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, et al. Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics 2011; 11:535-53.

35)   Ow SY, Salim M, Noirel J, Evans C, Rehman I, Wright PC. iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". J Proteome Res 2009; 8:5347-55.

36)   Schulze WX, Usadel B. Quantitation in mass-spectrometry-based proteomics. Annu Rev Plant Biol 2010; 61:491-516.

37)   Foster MW, Thompson JW, Que LG, Yang IV, Schwartz DA, Moseley MA, et al. Proteomic analysis of human bronchoalveolar lavage fluid after subsegemental exposure. J Proteome Res 2013; 12:2194-205.

38)   Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, et al. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. Eur Respir J 2015; 46:1308-21.

39)   Djukanovic R, Sterk PJ, Fahy JV, Hargreave FE. Standardised methodology of sputum induction and processing. Eur Respir J Suppl 2002; 37:1s-2s.

40)   Muntel, J., Fromion, V., Goelzer, A., Maaβ, S., Mäder, U., Büttner, K., Hecker, M. and Becher,

D., 2014. Comprehensive absolute quantification of the cytosolic proteome of Bacillus subtilis by data independent, parallel fragmentation in liquid chromatography/mass spectrometry (LC/MSE). Molecular & Cellular Proteomics, 13(4), pp.1008-1019

41) Silva JC, Denny R, Dorschel C, Gorenstein MV, Li GZ, Richardson K, et al. Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale. Mol Cell Proteomics 2006; 5:589-607.

42) Li GZ, Vissers JP, Silva JC, Golick D, Gorenstein MV, Geromanos SJ. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. Proteomics 2009; 9:1696-719.

43) Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007 Jan;8(1):118-27.

44) Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp DG, 2nd, et al. DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics 2008; 24:1556-8.

45) Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science 2015; 347:1260419.

46) Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. Mol Cell Proteomics 2011; 10:M110.006353.

47) Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics 2005; 5:3226-45.

48) Schenk S, Schoenhals GJ, de Souza G, Mann M. A high confidence, manually validated human blood plasma protein reference set. BMC Med Genomics 2008; 1:41.

49) Chen J, Ryu S, Gharib SA, Goodlett DR, Schnapp LM. Exploration of the normal human bronchoalveolar lavage fluid proteome. Proteomics Clin Appl 2008; 2:585-95.

50) Nguyen EV, Gharib SA, Palazzo SJ, Chow YH, Goodlett DR, Schnapp LM. Proteomic profiling of bronchoalveolar lavage fluid in critically ill patients with ventilator-associated pneumonia. PLoS One 2013; 8:e58782.

51) Nguyen EV, Gharib SA, Schnapp LM, Goodlett DR. Shotgun MS proteomic analysis of bronchoalveolar lavage fluid in normal subjects. Proteomics Clin Appl 2014; 8:737-47.

52) Wu J, Kobayashi M, Sousa EA, Liu W, Cai J, Goldman SJ, et al. Differential proteomic analysis of bronchoalveolar lavage fluid in asthmatics following segmental antigen challenge. Mol Cell Proteomics 2005; 4:1251-64.

53) Fumagalli M, Ferrari F, Luisetti M, Stolk J, Hiemstra PS, Capuano D, et al. Profiling the proteome of exhaled breath condensate in healthy smokers and COPD patients by LC-MS/MS. Int J Mol Sci 2012; 13:13894-910.

54) Muccilli V, Saletti R, Cunsolo V, Ho J, Gili E, Conte E, et al. Protein profile of exhaled breath condensate determined by high resolution mass spectrometry. J Pharm Biomed Anal 2015; 105:134-49.

55) Joo NS, Evans IA, Cho HJ, Park IH, Engelhardt JF, Wine JJ. Proteomic analysis of pure human airway gland mucus reveals a large component of protective proteins. PLoS One 2015; 10:e0116756.

56) Fang X, Yang L, Wang W, Song T, Lee CS, DeVoe DL, et al. Comparison of electrokinetics-based multidimensional separations coupled with electrospray ionization-tandem mass spectrometry for characterization of human salivary proteins. Anal Chem 2007; 79:5785-92.

57) Guo T, Rudnick PA, Wang W, Lee CS, Devoe DL, Balgley BM. Characterization of the human salivary proteome by capillary isoelectric focusing/nanoreversed-phase liquid chromatography coupled with ESI-tandem MS. J Proteome Res 2006; 5:1469-78.

58) Hu S, Xie Y, Ramachandran P, Ogorzalek Loo RR, Li Y, Loo JA, et al. Large-scale identification of proteins in human salivary proteome by liquid chromatography/mass spectrometry and two-dimensional gel electrophoresis-mass spectrometry. Proteomics 2005; 5:1714-28.

59) Ramachandran P, Boontheung P, Xie Y, Sondej M, Wong DT, Loo JA. Identification of N-linked glycoproteins in human saliva by

glycoprotein capture and mass spectrometry. J Proteome Res 2006; 5:1493-503.

60) Xie H, Rhodus NL, Griffin RJ, Carlis JV, Griffin TJ. A catalogue of human saliva proteins identified by free flow electrophoresis-based peptide separation and tandem mass spectrometry. Mol Cell Proteomics 2005; 4:1826-30.

61) Yan W, Apweiler R, Balgley BM, Boontheung P, Bundy JL, Cargile BJ, et al. Systematic comparison of the human saliva and plasma proteomes. Proteomics Clin Appl 2009; 3:116-34.

62) Cervero P, Himmel M, Kruger M, Linder S. Proteomic analysis of podosome fractions from macrophages reveals similarities to spreading initiation centres. Eur J Cell Biol 2012; 91:908-22.

63) Eligini S, Brioschi M, Fiorelli S, Tremoli E, Colli S, Banfi C. Data for proteomic analysis of Human monocyte-derived macrophages. Data Brief 2015; 4:177-9.

64) Acharya KR, Ackerman SJ. Eosinophil granule proteins: form and function. J Biol Chem 2014; 289:17406-15.

65) Kahn JE, Dutoit-Lefevre V, Duban-Deweer S, Chafey P, Pottiez G, Lefranc D, et al. Comparative proteomic analysis of blood eosinophils reveals redox signaling modifications in patients with FIP1L1-PDGFRA-associated chronic eosinophilic leukemia. J Proteome Res 2011; 10:1468-80.

66) Straub C, Burnham JP, White AC, Jr., Pazdrak K, Sanchez C, Watanabe LC, et al. Altered eosinophil proteome in a patient with hypereosinophilia from acute fascioliasis. Clin Vaccine Immunol 2011; 18:1999-2002.

67) Straub C, Pazdrak K, Young TW, Stafford SJ, Wu Z, Wiktorowicz JE, et al. Toward the Proteome of the Human Peripheral Blood Eosinophil. Proteomics Clin Appl 2009; 3:1151-73.

68) Yoon SW, Kim TY, Sung MH, Kim CJ, Poo H. Comparative proteomic analysis of peripheral blood eosinophils from healthy donors and atopic dermatitis patients with eosinophilia. Proteomics 2005; 5:1987-95.

69) Tomazella GG, da Silva I, Laure HJ, Rosa JC, Chammas R, Wiker HG, et al. Proteomic analysis of total cellular proteins of human neutrophils. Proteome Sci 2009; 7:32.

70) Tomazella GG, daSilva I, Thome CH, Greene LJ, Koehler CJ, Thiede B, et al. Analysis of detergent-insoluble and whole cell lysate fractions of resting neutrophils using high-resolution mass spectrometry. J Proteome Res 2010; 9:2030-6.

71) Trusch M, Tillack K, Kwiatkowski M, Bertsch A, Ahrends R, Kohlbacher O, et al. Displacement chromatography as first separating step in online two-dimensional liquid chromatography coupled to mass spectrometry analysis of a complex protein sample--the proteome of neutrophils. J Chromatogr A 2012; 1232:288-94.

72) Zhu J, Zhang H, Guo T, Li W, Li H, Zhu Y, et al. Quantitative proteomics reveals differential biological processes in healthy neonatal cord neutrophils and adult neutrophils. Proteomics 2014; 14:1688-97.

73) Brinkmann V, Reichard U, Goosmann C, Fauler B, Uhlemann Y, Weiss DS, et al. Neutrophil extracellular traps kill bacteria. Science 2004; 303:1532-5.

74) Guimaraes-Costa AB, Nascimento MT, Wardini AB, Pinto-da-Silva LH, Saraiva EM. ETosis: A Microbicidal Mechanism beyond Cell Death. J Parasitol Res 2012; 2012:929743.

75) Dalli J, Montero-Melendez T, Norling LV, Yin X, Hinds C, Haskard D, et al. Heterogeneity in neutrophil microparticles reveals distinct proteome and functional properties. Mol Cell Proteomics 2013; 12:2205-19.

76) Lominadze G, Powell DW, Luerman GC, Link AJ, Ward RA, McLeish KR. Proteomic analysis of human neutrophil granules. Mol Cell Proteomics 2005; 4:1503-21.

77) Rorvig S, Ostergaard O, Heegaard NH, Borregaard N. Proteome profiling of human neutrophil granule subsets, secretory vesicles, and cell membrane: correlation with transcriptome profiling of neutrophil precursors. J Leukoc Biol 2013; 94:711-21.

78) Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37:1-13.

79) Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4:44-57.

80) Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2012.

81)   Venny. An interactive tool for comparing lists with Venn's diagrams. 2007-2015.] Available from http://bioinfogp.cnb.csic.es/tools/venny/index.html.

82)   Bederson BB, Shneiderman B, Wattenberg M. Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies. In: Bederson BB, Shneiderman B, editors. The Craft of Information Visualization. San Francisco: Morgan Kaufmann; 2003. p. 257-78.

83)   Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 2015; 15:2597-601.

84)   Chen Y, Zhang Y, Yin Y, Gao G, Li S, Jiang Y, et al. SPD--a web-based secreted protein database. Nucleic Acids Res 2005; 33:D169-73.

85)   Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39:D561-8.

86)   Bateman A, Martin MJ, O'Donovan C, Magrane M, Apweiler R, Alpi E, et al. UniProt: a hub for protein information. Nucleic Acids Res 2015; 43:D204-12.

87)   Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 2009; 10:48.

88)   Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 2011; 6:e21800.

89)   Hinks TSC, Zhou X, Staples KJ, Dimitrov BD, Manta A, Petrossian T, Lum PY, Smith CG, Ward JA, Howarth PH, et al. Innate and adaptive T cells in asthmatic patients: relationship to severity and disease mechanisms. J Allergy Clin Immunol 2015;136:323–333.

90)   Li L, Cheng WY, Glicksberg BS, Gottesman O, Tamler R, Chen R, Bottinger EP, Dudley JT. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med 2015;7:311ra174.

91)   Nielson JL, Paquette J, Liu AW, Guandique CF, Tovar CA, Inoue T, Irvine KA, Gensel JC, Kloke J, Petrossian TC, et al. Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury. Nat Commun 2015;6:8581.

92)   Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, Carlsson G. Extracting insights from the shape of complex data using topology. Sci Rep 2013;3:1236.

93)   Carlsson G. Topology and data. Bull Am Math Soc 2009;46:255–308.

94)   Hoffman, G.E. and Schadt, E.E., 2016. variancePartition: interpreting drivers of variation in complex gene expression studies. BMC bioinformatics, 17(1), p.483.

95)   Ryckman C, Vandal K, Rouleau P, Talbot M, Tessier PA. Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. J Immunol 2003; 170:3233-42.

96)   Stockley RA. The multiple facets of alpha-1-antitrypsin. Ann Transl Med 2015; 3:130.

97)   Gaber F, Acevedo F, Delin I, Sundblad BM, Palmberg L, Larsson K, et al. Saliva is one likely source of leukotriene B4 in exhaled breath condensate. Eur Respir J 2006; 28:1229-35.

98)   Lee TH, Jang AS, Park JS, Kim TH, Choi YS, Shin HR, et al. Elevation of S100 calcium binding protein A9 in sputum of neutrophilic inflammation in severe uncontrolled asthma. Ann Allergy Asthma Immunol 2013; 111:268-75.e1.

99)   Gray RD, MacGregor G, Noble D, Imrie M, Dewar M, Boyd AC, et al. Sputum proteomics in inflammatory and suppurative respiratory diseases. Am J Respir Crit Care Med 2008; 178:444-52.

100)   Louten J, Mattson JD, Malinao MC, Li Y, Emson C, Vega F, et al. Biomarkers of disease and treatment in murine and cynomolgus models of chronic asthma. Biomark Insights 2012; 7:87-104.

101)   Zhao J, Zhu H, Wong CH, Leung KY, Wong WS. Increased lungkine and chitinase levels in allergic airway inflammation: a proteomics approach. Proteomics 2005; 5:2799-807.

102)   Bloemen K, Hooyberghs J, Desager K, Witters E, Schoeters G. Non-invasive biomarker sampling and analysis of the exhaled breath proteome. Proteomics Clin Appl 2009; 3:498-504.

103)   Goswami S, Angkasekwinai P, Shan M, Greenlee KJ, Barranco WT, Polikepahad S, et al. Divergent functions for airway epithelial matrix metalloproteinase 7 and retinoic acid in

experimental asthma. Nat Immunol 2009; 10:496-503.

104)   North ML, Khanna N, Marsden PA, Grasemann H, Scott JA. Functionally important role for arginase 1 in the airway hyperresponsiveness of asthma. Am J Physiol Lung Cell Mol Physiol 2009; 296:L911-20.

105)   Calvo FQ, Fillet M, de Seny D, Meuwis MA, Maree R, Crahay C, et al. Biomarker discovery in asthma-related inflammation and remodeling. Proteomics 2009; 9:2163-70.

106)   Jeong H, Rhim T, Ahn MH, Yoon PO, Kim SH, Chung IY, et al. Proteomic analysis of differently expressed proteins in a mouse model for allergic asthma. J Korean Med Sci 2005; 20:579-85.

107)   Chu HW, Thaikoottathil J, Rino JG, Zhang G, Wu Q, Moss T, et al. Function and regulation of SPLUNC1 protein in Mycoplasma infection and allergic inflammation. J Immunol 2007; 179:3995-4002.

108)   Kawakami M, Narumoto O, Matsuo Y, Horiguchi K, Horiguchi S, Yamashita N, et al. The role of CCR7 in allergic airway inflammation induced by house dust mite exposure. Cell Immunol 2012; 275:24-32.

109)   Brebner JA, Stockley RA. Recent advances in alpha-1-antitrypsin deficiency-related lung disease. Expert Rev Respir Med 2013; 7:213-29; quiz 30.

110)   Ishikawa N, Hattori N, Kohno N, Kobayashi A, Hayamizu T, Johnson M. Airway inflammation in Japanese COPD patients compared with smoking and nonsmoking controls. Int J Chron Obstruct Pulmon Dis 2015; 10:185-92.

111)   Ohlmeier S, Mazur W, Linja-Aho A, Louhelainen N, Ronty M, Toljamo T, et al. Sputum proteomics identifies elevated PIGR levels in smokers and mild-to-moderate COPD. J Proteome Res 2012; 11:599-608.

112)   Merkel D, Rist W, Seither P, Weith A, Lenter MC. Proteomic study of human bronchoalveolar lavage fluids from smokers with chronic obstructive pulmonary disease by combining surface-enhanced laser desorption/ionization-mass spectrometry profiling with mass spectrometric protein identification. Proteomics 2005; 5:2972-80.

113)   Shiratsuchi N, Asai K, Kanazawa H, Kyoh S, Tochino Y, Kodama T, et al. Measurement of soluble perforin, a marker of CD8+ T lymphocyte

activation in epithelial lining fluid. Respir Med 2011; 105:1885-90.

114)   Gohy ST, Detry BR, Lecocq M, Bouzin C, Weynand BA, Amatngalim GD, et al. Polymeric immunoglobulin receptor down-regulation in chronic obstructive pulmonary disease. Persistence in the cultured epithelium and role of transforming growth factor-beta. Am J Respir Crit Care Med 2014; 190:509-21.

115)   Stewart CE, Sayers I. Urokinase receptor orchestrates the plasminogen system in airway epithelial cell function. Lung 2013; 191:215-25.

116)   Calero C, Arellano E, Lopez-Villalobos JL, Sanchez-Lopez V, Moreno-Mata N, Lopez-Campos JL. Differential expression of C-reactive protein and serum amyloid A in different cell types in the lung tissue of chronic obstructive pulmonary disease patients. BMC Pulm Med 2014; 14:95.

117)   Ohlmeier S, Vuolanto M, Toljamo T, Vuopala K, Salmenkivi K, Myllarniemi M, et al. Proteomics of human lung tissue identifies surfactant protein A as a marker of chronic obstructive pulmonary disease. J Proteome Res 2008; 7:5125-32.

118)   Hara A, Sakamoto N, Ishimatsu Y, Kakugawa T, Nakashima S, Hara S, et al. S100A9 in BALF is a candidate biomarker of idiopathic pulmonary fibrosis. Respir Med 2012; 106:571-80.

119)   Landi C, Bargagli E, Carleo A, Bianchi L, Gagliardi A, Prasse A, et al. A system biology study of BALF from patients affected by idiopathic pulmonary fibrosis (IPF) and healthy controls. Proteomics Clin Appl 2014; 8:932-50.

120)   Mukae H, Ishimoto H, Yanagi S, Ishii H, Nakayama S, Ashitani J, et al. Elevated BALF concentrations of alpha- and beta-defensins in patients with pulmonary alveolar proteinosis. Respir Med 2007; 101:715-21.

121)   Jaffar J, Unger S, Corte TJ, Keller M, Wolters PJ, Richeldi L, et al. Fibulin-1 predicts disease progression in patients with idiopathic pulmonary fibrosis. Chest 2014; 146:1055-63.

122)   Bhargava M, Becker TL, Viken KJ, Jagtap PD, Dey S, Steinbach MS, et al. Proteomic profiles in acute respiratory distress syndrome differentiates survivors from non-survivors. PLoS One 2014; 9:e109713.

123)   Sloane AJ, Lindner RA, Prasad SS, Sebastian LT, Pedersen SK, Robinson M, et al. Proteomic analysis of sputum from adults and children with cystic fibrosis and from control

subjects. Am J Respir Crit Care Med 2005; 172:1416-26.

124)   Bullens DM, Truyen E, Coteur L, Dilissen E, Hellings PW, Dupont LJ, et al. IL-17 mRNA in sputum of asthmatic patients: linking T cell driven inflammation and granulocytic influx? Respir Res 2006; 7:135.

125)   Fahy JV, Kim KW, Liu J, Boushey HA. Prominent neutrophilic inflammation in sputum from subjects with asthma exacerbation. Journal of Allergy and Clinical Immunology 1995; 95:843-52.

126)   Cox G. Glucocorticoid treatment inhibits apoptosis in human neutrophils. Separation of survival and activation outcomes. J Immunol 1995; 154:4719-25.

127)   Dale DC, Fauci AS, Wolff SM. Alternate-day prednisone. Leukocyte kinetics and susceptibility to infections. N Engl J Med 1974; 291:1154-8.

128)   Pedersen B, Dahl R, Karlstrom R, Peterson CG, Venge P. Eosinophil and neutrophil activity in asthma in a one-year trial with inhaled budesonide. The impact of smoking. Am J Respir Crit Care Med 1996; 153:1519-29.

129)   Durdiakova J, Fabryova H, Koborova I, Ostatnikova D, Celec P. The effects of saliva collection, handling and storage on salivary testosterone measurement. Steroids 2013; 78:1325-31.

130)   Kamodyova N, Banasova L, Jansakova K, Koborova I, Tothova L, Stanko P, et al. Blood Contamination in Saliva: Impact on the Measurement of Salivary Oxidative Stress Markers. Dis Markers 2015; 2015:479251.

131)   Lőrincz, Á.M., Schütte, M., Timár, C.I., Veres, D.S., Kittel, Á., McLeish, K.R., Merchant, M.L. and Ligeti, E., 2015. Functionally and morphologically distinct populations of extracellular vesicles produced by human neutrophilic granulocytes. Journal of leukocyte biology, 98(4), pp.583-589.

132)   Kobayashi K, Ogata H, Morikawa M, Iijima S, Harada N, Yoshida T, et al. Distribution and partial characterisation of IgG Fc binding protein in various mucin producing cells and body fluids. Gut 2002; 51:169-76.

133)   Thornton DJ, Sheehan JK. From mucins to mucus: toward a more coherent understanding of this essential barrier. Proc Am Thorac Soc 2004; 1:54-61.

134)   Andersson C, Zaman MM, Jones AB, Freedman SD. Alterations in immune response and PPAR/LXR regulation in cystic fibrosis macrophages. Journal of Cystic Fibrosis 2008; 7:68-78.

135)   Hong C, Walczak R, Dhamko H, Bradley MN, Marathe C, Boyadjian R, et al. Constitutive activation of LXR in macrophages regulates metabolic and inflammatory gene expression: identification of ARL7 as a direct target. J Lipid Res 2011; 52:531-9.

136)   Renga B, Migliorati M, Mencarelli A, Fiorucci S. Reciprocal regulation of the bile acid-activated receptor FXR and the interferon-γ-STAT-1 pathway in macrophages. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 2009; 1792:564-73.

137)   Stojancevic M, Stankov K, Mikov M. The impact of farnesoid X receptor activation on intestinal permeability in inflammatory bowel disease. Can J Gastroenterol 2012; 26:631-7.

138)   Birrell MA, De Alba J, Catley MC, Hardaker E, Wong S, Collins M, et al. Liver X receptor agonists increase airway reactivity in a model of asthma via increasing airway smooth muscle growth. J Immunol 2008; 181:4265-71.

139)   Castrillo A, Joseph SB, Vaidya SA, Haberland M, Fogelman AM, Cheng G, et al. Crosstalk between LXR and Toll-like Receptor Signaling Mediates Bacterial and Viral Antagonism of Cholesterol Metabolism. Molecular Cell 2003; 12:805-16.

140)   Tall AR, Yvan-Charvet L. Cholesterol, inflammation and innate immunity. Nat Rev Immunol 2015; 15:104-16.

141)   Bezemer GF, Sagar S, van Bergenhenegouwen J, Georgiou NA, Garssen J, Kraneveld AD, et al. Dual role of Toll-like receptors in asthma and chronic obstructive pulmonary disease. Pharmacol Rev 2012; 64:337-58.

142)   Phipps S, Lam CE, Foster PS, Matthaei KI. The contribution of toll-like receptors to the pathogenesis of asthma. Immunol Cell Biol 2007; 85:463-70.

143)   Chiba Y, Misawa M. MicroRNAs and their therapeutic potential for human diseases: MiR-133a and bronchial smooth muscle hyperresponsiveness in asthma. J Pharmacol Sci 2010; 114:264-8.

144)   Yoshii A, Iizuka K, Dobashi K, Horie T, Harada T, Nakazawa T, et al. Relaxation of contracted rabbit tracheal and human bronchial

smooth muscle by Y-27632 through inhibition of Ca2+ sensitization. Am J Respir Cell Mol Biol 1999; 20:1190-200.

145)   Bastarache, J. A., Fremont, R. D., Kropski, J. A., Bossert, F. R., & Ware, L. B. (2009). Procoagulant alveolar microparticles in the lungs of patients with acute respiratory distress syndrome. AJP: Lung Cellular and Molecular Physiology, 297(6), L1035–L1041.

146)   Sadallah, S., Eken, C., & Schifferli, J. A. (2011). Ectosomes as modulators of inflammation and immunity. *Clinical & Experimental Immunology*, *163*(1), 26–32. https://doi.org/10.1111/j.1365-2249.2010.04271.x

147)   Pick E, Gorzalczany Y, Engel S. Role of the rac1 p21-GDP-dissociation inhibitor for rho heterodimer in the activation of the superoxide-forming NADPH oxidase of macrophages. Eur J Biochem 1993; 217:441-55.

148)   Robbins PD, Morelli AE. Regulation of immune responses by extracellular vesicles. Nat Rev Immunol 2014; 14:195-208.

149)   Hess C, Sadallah S, Hefti A, Landmann R, Schifferli JA. Ectosomes released by human neutrophils are specialized functional units. J Immunol 1999; 163:4564-73.

150)   Admyre C, Telemo E, Almqvist N, Lotvall J, Lahesmaa R, Scheynius A, et al. Exosomes - nanovesicles with possible roles in allergic inflammation. Allergy 2008; 63:404-8.

151)   Spencer LA, Bonjour K, Melo RC, Weller PF. Eosinophil secretion of granule-derived cytokines. Front Immunol 2014; 5:496.

152)   Torregrosa Paredes P, Esser J, Admyre C, Nord M, Rahman QK, Lukic A, et al. Bronchoalveolar lavage fluid exosomes contribute to cytokine and leukotriene production in allergic asthma. Allergy 2012; 67:911-9.

Table of Content Graphic