ERJ open research
THE BEST IN OPEN ACCESS BASIC, TRANSLATIONAL & CLINICAL RESPIRATORY RESEARCH

Early View

Original research article

Multiomics links global surfactant dysregulation with airflow obstruction and emphysema in COPD

Ventzislava A. Hristova, Alastair Watson, Raghothama Chaerkady, Matthew S. Glover, Jodie Ackland, Bastian Angerman, Graham Belfield, Maria G. Belvisi, Hannah Burke, Doriana Cellura, Howard W. Clark, Damla Etal, Anna Freeman, Ashley I Heinson, Sonja Hess, Michael Hühn, Emily Hall, Alex Mackay, Jens Madsen, Christopher McCrae, Daniel Muthas, Steven Novick, Kristoffer Ostridge, Lisa Öberg, Adam Platt, Anthony D. Postle, C. Mirella Spalluto, Outi Vaarala, Junmin Wang, Karl J. Staples, Tom M.A Wilkinson, on behalf of the MICA II Study group

Please cite this article as: Hristova VA, Watson A, Chaerkady R, *et al*. Multiomics links global surfactant dysregulation with airflow obstruction and emphysema in COPD. *ERJ Open Res* 2022; in press (https://doi.org/10.1183/23120541.00378-2022).

Multiomics links global surfactant dysregulation with airflow obstruction and emphysema in COPD

Authors:

Ventzislava A. Hristova[1*], Alastair Watson[2,3*], Raghothama Chaerkady[1], Matthew S. Glover[1], Jodie Ackland[2], Bastian Angerman[4], Graham Belfield[5], Maria G. Belvisi[6,7], Hannah Burke[2,8], Doriana Cellura[2], Howard W. Clark[9,10], Damla Etal[5], Anna Freeman[2,8], Ashley I Heinson[2], Sonja Hess[1], Michael Hühn[4], Emily Hall[2], Alex Mackay[4,7], Jens Madsen [2], Christopher McCrae[11], Daniel Muthas[4], Steven Novick[12], Kristoffer Ostridge[2,4], Lisa Öberg[4], Adam Platt[13], Anthony D. Postle[2], C. Mirella Spalluto[2], Outi Vaarala[14], Junmin Wang[1], Karl J. Staples[2,8], Tom M.A Wilkinson[2,8 #] on behalf of the MICA II Study group[&]

*These authors contributed equally to this work

#Corresponding author:

Professor Tom Wilkinson

Faculty of Medicine, University of Southampton, Southampton, UK

t.wilkinson@soton.ac.uk

[1]Dynamic Omics, Centre for Genomics Research (CGR), Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, USA

[2]Faculty of Medicine, University of Southampton, Southampton, UK

[3]School of Clinical Medicine, University of Cambridge, Cambridge, UK

[4]Translational Science and Experimental Medicine, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[5]Translational Genomics, Discovery Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[6]Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[7]NHLI, Imperial College, London, UK

[8]NIHR Southampton Biomedical Research Centre, University Hospital, Southampton, UK

[9]Neonatology, Faculty of Population Health Sciences, EGA Institute for Women's Health, University College London, London, UK

[10]NIHR University College London Hospital Biomedical Research, University College London Hospital, London, UK

[11]Translational Science and Experimental Medicine, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, USA

[12]Data Sciences & Quantitative Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, USA

[13]Translational Science and Experimental Medicine, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

[14]Faculty of Medicine, University of Helsinki, Helsinki, Finland

[&]For MICAII study group see end of manuscript

Text word count: 3,262

Author's contributions

Ventzislava A. Hristova analysed proteomics, lipidomics and subject data. VH wrote the manuscript with input from all authors; approved the final version of the manuscript. Alastair Watson conceptualized parts of the project; contributed to methodology, sample collection and processing and project administration; wrote the original draft with input from all authors and approved the final version of the manuscript. Raghothama Chaerkady designed, conducted and analysed the proteomics experiments. RC contributed to writing of the manuscript; approved the final version of the manuscript. Matthew S. Glover designed, conducted and analysed the lipidomics experiments. MG contributed to writing of the manuscript; approved the final version of the manuscript. Jodie Ackland analysis, helped write the original draft, reviewing and editing of manuscript; approved the final version of the manuscript. Bastian Angerman analysed data, contributed to writing of the manuscript, approved the final version of the manuscript. Graham Belfield contributed to the design, optimisation and analysis of the transcriptomic experiments, contributed to writing of the manuscript, approved the final version of the manuscript. Maria G. Belvisi conceptualized parts of the project, provided resources, contributed to writing of the manuscript, approved the final version of the manuscript. Hannah Burke sample collection methodology, sample collection, contributed to the original draft, approved the final version of the manuscript. Doriana Cellura sample collection and processing, approved the final version of the manuscript. Howard Clark, conceptualization of parts of the project; editing and approved the final version of the manuscript. Damla Etal sample post processing methodology and optimization, conducted the transcriptomic experiments, contributed to writing of the

manuscript, approved the final version of the manuscript. Anna Freeman Sample collection methodology, sample collection, contributed to the original draft, approved the final version of the manuscript. Emily Hall, ELISA analysis; editing and approved the final version of the manuscript. Ashley I Heinson contributed to writing of the manuscript, approved the final version of the manuscript. Sonja Hess designed the proteomics and lipidomics experiments and analysed the data. SH curated subject and medication metadata for subsequent data and statistical analysis. SH oversaw the omics studies and provided resources. SH contributed to writing of the manuscript; approved the final version of the manuscript. Michael Hühn conceptualized parts of the project, analysis, contributed to writing of the manuscript, approved the final version of the manuscript. Alex Mackay conceptualized parts of the project, analysis, supervised the project, contributed to writing of the manuscript, approved the final version of the manuscript. Jens Madsen, conceptualization of parts of the project; editing and approved the final version of the manuscript. Christopher McCrae conceptualized parts of the project, analysis, contributed to writing of the manuscript, approved the final version of the manuscript. Daniel Muthas conceptualized parts of the project, analysis, contributed to writing of the manuscript, approved the final version of the manuscript. Kristoffer Ostridge conceptualized the project and designed the study, provided resources, supervised the project, sample collection methodology, sample collection, contributed to writing of the manuscript, approved the final version of the manuscript. Lisa Öberg analysed data, contributed to writing of the manuscript, approved the final version of the manuscript. Adam Platt conceptualized parts of the project, provided resources, contributed to writing of the manuscript, approved the final version of the manuscript. Anthony D. Postle contributed to the original draft, approved the final version of the manuscript. C. Mirella Spalluto Sample collection methodology, sample collection and

processing, project administration, supervised the project, contributed to writing of the manuscript, approved the final version of the manuscript. Outi Vaarala conceptualized parts of the project, contributed to writing of the manuscript, approved the final version of the manuscript. Junmin Wang and Steven Novick analysed proteomics and lipidomics data. JW contributed to writing of the manuscript; approved the final version of the manuscript. Karl J. Staples conceptualized the project, performed sample collection methodology, sample collection and processing, project administration, supervised the project and contributed to the original draft, approved the final version of the manuscript. Tom Wilkinson conceptualized the project and designed the study, provided resources, acquired funding, supervised the project, contributed to writing of the manuscript, approved the final version of the manuscript.

Funding

The study was funded by AstraZeneca. AstraZeneca reviewed the publication, without influencing the opinions of the authors, to ensure medical and scientific accuracy, and the protection of intellectual property. The clinical study was supported by the NIHR Southampton BRC. The corresponding author had access to all data in the study and had the final responsibility for the decision to submit the manuscript for publication.

Running title: Multiomics: Surfactant dysregulation in COPD

Take home message: Multiomics demonstrates global surfactant dysregulation in COPD, associating with emphysema and airway obstruction severity. These findings frame the need for future studies to explore the potential for novel surfactant-targeting therapeutics.

Abstract

Rationale

Pulmonary surfactant is vital for lung homeostasis as it reduces surface tension to prevent alveolar collapse and provides essential immune-regulatory and anti-pathogenic functions. Previous studies demonstrated dysregulation of some individual surfactant components in COPD.

Objectives

We investigated relationships between COPD disease measures and dysregulation of surfactant components to gain new insights about potential disease mechanisms.

Methods

Bronchoalveolar lavage proteome and lipidome were characterised in ex-smoking mild/moderate COPD subjects (n=26) and healthy ex-smoking (n=20) and never-smoking (n=16) controls using mass spectrometry. Serum surfactant protein analysis was performed.

Results

Total phosphatidylcholine, phosphatidylglycerol, phosphatidylinositol and surfactant protein (SP)-B, SP-A and SP-D concentrations were lower, COPD *vs*. controls, log2 fold change (log2FC)=-2.0, -2.2, -1.5, -0.5, -0.7, -0.5 (adj. p-value<0.02), respectively, and correlated with lung function. Total phosphatidylcholine, phosphatidylglycerol, phosphatidylinositol and SP-A, SP-B, SP-D, NAPSA and CD44 inversely correlated with CT small airways disease measures (E/I MLD), r=-0.56, r=-0.58, r=-0.45, r=-0.36, r=-0.44, r=-0.37, r=-0.40, r=-0.39 (adj. p-value<0.05). Total phosphatidylcholine, phosphatidylglycerol, phosphatidylinositol and SP-A, SP-B, SP-D and NAPSA inversely correlated with emphysema (%LAA): r=-0.55, r=-0.61, r=-

0.48, r=-0.51, r=-0.41, r=-0.31, r=-0.34, respectively (adj. p-value<0.05). Neutrophil elastase, known to degrade SP-A and SP-D, was elevated, COPD *vs.* controls, log2FC of 0.40 (adj. p-value=0.0390) and inversely correlated with SP-A and SP-D. Serum SP-D was increased in COPD *vs.* HV-ES, and predicted COPD status, AUC=0.85.

Conclusions

Using a multiomics approach we, for the first time, demonstrate global surfactant dysregulation in COPD which was associated with emphysema giving new insights about potential mechanisms underlying the cause or consequence of disease.

Keywords: COPD, Pulmonary Surfactant, Multiomics, Proteomics, Lipidomics

Abstract word count: 250

Introduction

Chronic obstructive pulmonary disease (COPD) is a leading cause of morbidity and mortality worldwide. There is still much to be understood of the mechanistic processes underlying its pathology as it drives such an important unmet clinical need [1-4]. Pulmonary surfactant homeostasis is critical to healthy lung function, as it coats the air-liquid interface reducing surface tension and preventing alveolar collapse at end-expiration [5,6]. Tightly controlled synthesis, secretion and subsequent recycling of surfactant are key to facilitate these essential functions. Thus, emphysematous changes in COPD and loss of alveolar type II (AT2) cells, which produce surfactant, may lead to disrupted surfactant synthesis and homeostasis and requires study.

Pulmonary surfactant is comprised of ~90% lipids, and 10% proteins. Phosphatidylcholine (PC) accounts for >80% of surfactant lipids, with phosphatidylglycerol (PG) for ~15%, and the remainder is phosphatidylethanolamine (PE), phosphatidylinositol (PI), sphingomyelin and other lipids [7-10]. Surfactant proteins (SP)-B and SP-C are small hydrophobic proteins, with essential biophysical roles in surfactant packaging, recycling, and maintaining surfactant structure [5,11]. SP-B is essential for reducing surface tension and its production is regulated by napsin-A [11,12]. SP-C regulation is not fully understood, but SP-B, Cathepsin H (CTSH) and Nedd4 have been suggested to facilitate its production [13-15]. In contrast, SP-A, composed of SP-A1 and SP-A2, and SP-D, are large, soluble, innate immune defence molecules with essential immunomodulatory and homeostatic lung functions [16-20]. These prevent infection and help clear bacterial, viral and fungal pathogens, whilst preventing aberrant inflammation and damage to the delicate epithelial-endothelial barrier [19-22].

Surfactant dysregulation may play a role in pathological processes underlying COPD through changes in alveolar tension and development of emphysema [23]. In addition, SP-D, specifically, has long been known to be deficient in COPD, which may predispose to both exacerbations and inflammatory processes [24]. However, there are contradicting reports around SP-A pulmonary levels in COPD and levels of SP-B and SP-C remain to be fully elucidated [25-27]. A recent study reported surfactant lipids to be in lower abundance in a small cohort of COPD subjects as compared with non-smoking controls [28], however, the impact of disease on surfactant in the absence of current smoking remains to be elucidated. We used an unbiased comprehensive multiomics approach to characterise proteome and lipidome differences in bronchoalveolar lavage fluid (BAL) in well-characterised COPD subjects and healthy ex-smoking controls to better understand surfactant dysregulation in COPD and glean insights about potential mechanisms underlying the cause or consequence of disease.

Methods

Subjects

The MICAII study recruited subjects with mild or moderate COPD (GOLD guidelines), alongside healthy ex-smoking volunteers (HV-ES), all had ≥10-pack year history, but had stopped smoking ≥6 months prior to enrolment [29-32]. Healthy volunteer never-smokers (HV-NS) were also recruited. All MICAII study subjects with recovered BAL supernatants suitable for proteomic and lipidomic analysis were included in this study (demographics given in Table 1). Total subject numbers per group, therefore, differ slightly from previous publications on the MICAII study. Further details about this cohort have previously been reported [29-32]. Matched serum was also used for complimentary proteomic analysis. This included both participants within the main cohort and some additional participants who were removed from the study prior to bronchoscopy due to numerous reasons, including subject request, not being suitable for bronchoscopy, or not fitting the inclusion criteria as set out in the methodology (demographics given in Table S1). Subjects were recruited from a combination of sources, including established research databases held within the University hospital Southampton, contact by clinicians involved or aware of the study within the hospital and local health care facilities and through subjects responding to study adverts/posters. All subjects gave written informed consent. The study was approved by National Research Ethics Service South Central – Hampshire A and Oxford C Committees (LREC no: 15/SC/0528).

As previously described, all subjects underwent volumetric CT chest scans in full inspiration and maximum expiration using a Siemens Sensation 64 scanner [33]. Low attenuation area below -950 Hounsfield Units (%LAA) was calculated as a measure of emphysema and

prebronchodilator, single-breath diffusion was performed, as per guidelines, with percent predicted carbon monoxide transfer coefficient calculated (TLCO%). A surrogate marker for small airways disease was measured using the ratio of mean lung attenuation on expiratory and inspiratory scans (E/I MLD).

Sample Collection

Sampling was undertaken using fibre optic bronchoscopy, and BAL was recovered and processed as previously described [29,30]. Macrophages were sorted by flow cytometry using forward scatter width (FSC-W) and forward scatter area (FSC-A) and subsequently CD45, CD163 and HLA-DR expression. Serum was isolated from blood as previously described[34].

Experimental Design of Analysis

The three groups and other statistically-modeled covariates (e.g., age, gender) were balanced through statistical D-optimal block design via the optBlock() function in the AlgDesign library in R. Each TMT 11-plex contained all three groups and balance was achieved across the 10 plexes.

Proteomics and Lipidomics

BAL supernatants were processed using an S-Trap-based method (protifi.com). Proteins were digested with trypsin/lysC (Promega, Madison, WI). Resulting peptides were desalted and subjected to tandem mass tag (TMT) (cat. No. A34808, Thermo Fisher Scientific) labelling for 11plex-TMT analysis, according to manufacturer's instructions. LC-MS/MS analysis was carried out on a Q Exactive HF-X (Thermo Fisher Scientific) mass spectrometer interfaced with a Dionex 3000 RSLCnano (acquisition parameters outlined in supplementary methods). Data analysis was undertaken with Proteome Discoverer 2.3 (Thermo Fisher

Scientific) and Mascot (version 2.6.0) using the latest Uniprot Human protein database (search parameters outlined in supplementary methods). Total protein levels were consistent across BAL samples, no outliers were observed before normalization. Protein quantitation was analysed using Perseus software v 1.6.15.0. and protein abundances were normalized to total protein levels [35,36].

Serum was depleted with HighSelect Top14 Abundant Protein Depletion Resin and digested using EasyPep 96 MS Prep Kit (Thermo Fisher Scientific), according to manufacturer's instructions. Serum analysis was undertaken by data-independent acquisition (DIA) on an Exploris 480 mass spectrometer interfaced with a Dionex 3000 RSLCnano (Thermo Fisher Scientific), DIA analysis was performed with Spectronaut v15 (Biognosys) (sample processing, acquisition and analysis parameters outlined in supplementary methods).

Lipid extraction from BAL supernatants were performed using a modified Maytash method [37]. LC-MS/MS analysis was performed on a Vanquish UHPLC – Orbitrap ID-X Tribrid MS (Thermo Scientific). Lipidomic data was analysed using MS-DIAL version 4 [38]. Detailed parameters of lipidomic experiments performed by LC-MS/MS and data analysis are outlined in supplementary methods.

SP-D ELISA was performed to confirm mass spectrometry results using a rabbit polyclonal anti-recombinant fragment of human SP-D (anti-rfhSP-D) antibody capture antibody and biotinylated mouse anti-human SP-D (Hyb246-04) detection antibody, with streptavidin-horse radish peroxidase [39]. Quantification was through comparison with a recombinant full length human SP-D standard [40].

Bioinformatic and Statistical Analysis

We fitted a linear model separately for each lipid and lipid class, and a linear mixed model for each protein, to understand differences between COPD and control cohorts, while accounting for effects of age, gender, and experimental design. For participants with two BAL samples, the average of the two samples per subject was taken prior to constructing and fitting the models. The model for any lipid was specified as:

$$y_{jk} = \mu_j + \beta_1 \times Age_k + \beta_2 \times Sex_k + e_{jk}$$

where y_{jk} was the log2-transformed abundance or composition of the lipid, the jth group (COPD, HV-ES, HV-NS), and the k^{th} subject (of a certain age and sex). $e_{jk} \sim N(0, \sigma_e^2)$ denoted subject-to-subject variability. The model for a lipid class had the same formula, except that y_{jk} was the log2-transformed summed abundance of the lipids belonging to that class.

For BAL proteomics, TMT batches were added as a random effect. Two samples of the same subject were placed in the same TMT batch except for one subject, which was excluded from model fitting. The model for any protein was specified as:

$$y_{jkl} = \mu_j + \beta_1 \times Age_k + \beta_2 \times Sex_k + T_l + e_{jkl}$$

where all terms were defined the same as above, but with $T_l \sim N(0, \sigma_T^2)$ and $e_{jkl} \sim N(0, \sigma_e^2)$ respectively denoting TMT batch and subject-to-subject variability.

The model coefficients were estimated using the lmer() function in the lme4 R package [41] (R version 3.6.0) [13,14]. Pairwise comparisons of estimated marginal means were conducted using the emmeans R package. The abundance of a protein/lipid, lipid composition, or summed abundance of a lipid class was considered significantly different between two groups if its adjusted p-value (p-value adjusted via Benjamini-Hochberg false-discovery-rate

(FDR) method) was <0.05. Box plots were made after effects of age, gender and random effects, if present, were subtracted from the model fits. The Spearman's rank correlation test was conducted to determine the association between two variables (p-value adjusted via FDR method). For participants with two BAL samples, the average of the two samples was taken prior to the correlation analysis.

A separate analysis was undertaken to evaluate the predictive value of serum surfactant proteins for use as a potential COPD biomarker. This was done through development of a logistic regression model that classifies COPD status based on serum SP-D. The model was trained using data from twenty-nine donor-matched serum samples for which serum SP-D was detected.

Transcriptomics

For subjects where sufficient alveolar macrophages were able to be purified, total RNA was extracted from BAL fluid purified macrophages using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen), as previously described[32]. Gene expression profiles of isolated macrophages from 47 subjects (15 HV-NS, 18 HV-ES and 14 COPD; demographics given in Table S3) were assessed using total RNA-sequencing on an Illumina NovaSeq 6000 platform as outlined in the supplementary methods and as previously described [32]. To explore surfactant transcriptomic differences in macrophages, differential gene expression analysis was performed with DESeq2 (v1.26.0) and weighted gene correlation network analysis (WGCNA) [42] was also implemented (full details in supplementary methods).

Results

Subject demographics

BAL analysis was undertaken in 26 COPD subjects, 20 HV-ES and 16 HV-NS, clinical characteristics are summarised in Table 1. There were no significant differences between HV-NS and HV-ES in age, sex, BMI or lung function (FEV1 and FEV1/FVC), but expected differences were seen in pack year history. There were no significant differences between HV-ES and COPD volunteers in age, sex or BMI. However, there were expected significant differences in lung function, CT and physiological measures of emphysema and small airways disease. There were no significant differences in proportion of eosinophils, macrophages or neutrophils between HV-NS and HV-ES or HV-ES and COPD BAL. There was a significantly lower median proportion of lymphocytes in BAL in COPD *vs.* HV-ES (1.02 *vs.* 0.03, p=0.034) (Table 2).

Surfactant lipid dysregulation in COPD

There were no significant differences in BAL lipid composition between males and females (Figure S1A). No differences were seen in total PC, PG and PI lipids or surfactant-specific PC and PG species between HV-ES and HV-NS (Figure 1A). However, a lower total concentration of PC, PG and PI were observed in COPD compared to HV-ES, log2FC of -2.0, -2.2 and -1.5 (all adj. p-value<0.0001), respectively (Figure 1B). Concentrations of PC, PG and PI were further decreased when compared with HV-NS (Figure 1B).

To understand specific BAL surfactant phospholipid differences in COPD, we performed a detailed examination of distinct PC and PG molecular species abundances as a percentage of each phospholipid type. Lower concentrations of desaturated PC 32:0, corresponding to

dipalmitoylphosphatidylcholine (DPPC), were observed in COPD *vs*. HV-ES (log2 FC -2.5, adj. p-value<0.0001). As a percentage of total PC, PC 32:0 was significantly lower in COPD *vs*. HV-ES BAL, log2FC of -0.47 (adj. p-value=0.001) (Figure 1C). As a percentage of total PC abundance, PC 30:0 was not significantly lower in COPD BAL, log2FC -0.16 (adj. p-value=0.4080), while PC 34:1 was a higher percentage of total PC in COPD subjects compared to HV-ES, log2FC of 0.28 (adj. p-value=0.0244) (Figure 1C). PG 34:1 and 36:2, as a percentage of total PG, were not significantly different between COPD and HV-ES subject BAL, with log2FC of –0.06 (adj. p-value=0.6433) and log2FC of –0.03 (adj. p-value=0.8709), respectively (Figure 1D). The percentage of PG 36:1 out of total PG was significantly lower in COPD *vs.* HV-ES BAL with log2FC –0.29 (adj. p-value=0.0322) (Figure 1D).

BAL Ceramides CER 42:1, CER 40:1, CER 34:1, CER 42:2, CER 38:1 and CER 36:1 were significantly higher in COPD compared to HV-NS, with log2FC of 0.6, 0.9, 1.2, 0.8, 1.2 and 1.1, respectively (all adj. p-value<0.05). Other sphingolipids were not significantly different. No differences in sphingolipids were seen between COPD and HV-ES.

Correlation of lipid surfactant with lung disease measures

We next explored if BAL surfactant dysregulation associated with markers of lung function, CT and physiological measures of emphysema (high %LAA and low TLCO%, respectively) and a surrogate CT marker of small airways disease (E/I MLD). Low BAL surfactant levels were seen in subjects with worse lung function. Total PC, PG and PI BAL levels correlated with both FEV1 (r=0.52, r=0.56 and r=0.47 respectively, all adj. p-value<0.01) and FEV1/FVC (r=0.68, r=0.73 and r=0.60 respectively, all adj. p-value<0.01) (Figure 2). Low BAL surfactant concentrations were seen in subjects with the most emphysema. Total BAL PC, PG and PI correlated with TLCO%, r=0.41, r=0.42, r=0.39, respectively (all adj. p-value<0.01), and

inversely correlated with %LAA, r=-0.55, r=-0.61, r=-0.48, respectively (all adj. p-value<0.01). E/I MLD negatively correlated with total PC (r=-0.56), total PG (r=-0.58) and total PI (r=-0.45), all adj. p-value<0.01.

Surfactant protein dysregulation in COPD

There were no significant differences in male *vs.* female BAL proteomes (Figure S1B). No significant proteome differences, including in pulmonary surfactant protein levels, were observed between HV-ES and HV-NS (Figure 3A). In contrast, SP-B, SP-A and SP-D BAL levels were significantly lower in COPD *vs.* HV-ES, log2FC of -0.5, -0.7 and -0.5 (all adj. p-value<0.02), respectively (Figure 3B). Napsin A and Pro-Cathepsin H, proteins important in SP-B synthesis[11,12], were lower in COPD *vs.* HV-ES BAL, with log2FC of -0.6 (adj. p-value=0.0017) and -0.4 (adj. p-value=0.0466), respectively. In BAL SP-B correlated with NAPSA abundance (r=0.72, $p<2.2e^{-16}$) and CTSH levels (r=0.5, $p=4.8e^{-5}$) (Figure S2A and S2B). CD44 antigen, which has been reported to play a role in surfactant homeostasis [43,44], was lower in COPD *vs.* HV-ES BAL with a log2FC of -0.6 (adj. p-value=0.0025) (Figure 3B).

Neutrophil elastase, an enzyme known to degrade SP-A and SP-D [45-47], was elevated in COPD *vs.* HV-ES BAL, log2FC of 0.40 (adj. p-value=0.0390) (Figure 3B). Furthermore, concentrations of neutrophil elastase inversely correlated with concentrations of SP-A, r=-0.36 (p=0.0046), and SP-D, r=-0.33 (p=0.0083) (Figure S2 C,D). No other proteases were significantly different in concentration between COPD and HV-ES BAL.

To validate our findings, we subsequently confirmed our mass spectrometry data of lower BAL SP-D in COPD using ELISA. As expected, SP-D concentrations were signficantly lower in COPD *vs.* HV-ES, median concentration of 1878.90 ng/ml *vs.* 4684.92 ng/ml (p<0.01) (Figure 3C). There was no difference in SP-D concentration in BAL between HV-NS *vs.* HV-ES.

Correlation of surfactant proteins with lung disease measures

SP-A, SP-B, SP-D and CD44 BAL levels correlated with FEV1 (r=0.50, r=0.46, r=0.29 and r=0.41, all adj. p-value<0.05). SP-A, SP-B, SP-D and NAPSA BAL levels also correlated with FEV1/FVC (r=0.77, r=0.68, r=0.51 and r=0.52, respectively, all adj. p-value<0.01) (Figure 2). BAL SP-A and SP-B correlated with TLCO%, r=0.45 and r=0.37, respectively (both adj. p-value<0.01). Furthermore, BAL levels of SP-A, SP-B, SP-D and NAPSA inversely correlated with %LAA, r=-0.51, r=-0.41, r=-0.31 and r=-0.34, respectively (all adj. p value<0.05). Finally, low surfactant protein abundance in BAL was associated with a CT measure of small airways disease; E/I MLD negatively correlated with SP-A (r=-0.36), SP-B (r=-0.44), SP-D (r=-0.37), NAPSA (r=-0.40) and CD44 (r=-0.39), all p-values <0.01.

Surfactant protein differences in serum, potential of SP-D as a biomarker

To understand if dysregulation of surfactant could be detected in blood, we undertook proteomic analysis of serum samples. Serum analysis were performed in 35 COPD subjects, 22 HV-ES and 19 HV-NS from the MICAII cohort (demographics for included participants given in Table S1 and serum protein identifications in Table S2). Serum SP-D was significantly higher in COPD *vs.* HV-ES, mean SP-D abundance intensities of ~5000 and ~3000, respectively (p-value = 0.0095). Significantly more serum SP-D was present in COPD *vs*. HV-NS, mean abundance intensities of ~5000 and ~2450, respectively (p= 0.0030) (Figure 4A). SP-D levels in circulation negatively correlated with BAL SP-D abundance in donor-matched samples (Spearman's rank correlation r=-0.37, p-value = 0.05) (Figure 4B). Serum SP-B was detected in six COPD subjects and two HV-ES, but not in HV-NS (Figure 4C). Serum CTSH was detected in thirty donors, but did not show differential abundance in COPD *vs.* controls. Serum CTSH levels did not correlate with BAL CTSH abundance (r=-0.15, p=0.49) (Figure 4D).

Notably, serum SP-D levels were able to predict COPD status by the logistic regression model with an area under the curve (AUC) of 0.85 (Figure 4E).

Alveolar macrophage gene expression

A recent study in *Cd44-/-* knockout mice, demonstrated that deficiency of CD44 on alveolar macrophages, disrupted surfactant lipid homeostasis [44]. We analysed the transcriptome of purified BAL macrophages (demographics for included participants given in Table S3) and found no significant activation of transcriptomic pathways involved in macrophage lipid turnover or potentially involved in surfactant metabolism using either differentially expressed gene (DEG) analysis or WGNCA (Figure S3) [48].

Discussion

This study delineates the BAL proteome and lipidome of a well-characterized cohort of mild-to-moderate COPD subjects and ex-smoking and never-smoking controls. Using multiomics we gained a comprehensive understanding of surfactant dysregulation in COPD, independent of current smoking effects, to glean insights about potential explanatory mechanisms [49].

We report lower concentrations of surfactant lipids, surfactant proteins and proteins involved in surfactant synthesis in BAL from COPD subjects *vs.* controls, which correlated with airflow obstruction. Furthermore, we demonstrate an association with emphysema, highlighting that decreased surfactant concentrations could be driving mechanisms underlying this pathology or be a consequence of emphysematous changes, or both. Network modelling previously suggested lung surface tension to be important in emphysema pathophysiology through its influence on lung recoil [23]. Our study adds to these findings by demonstrating that surfactant, the key lung surface tension regulator, is decreased in COPD *vs.* HV-ES. This expands on prior small studies reporting lower surfactant lipids in BAL from COPD *vs.* healthy non-smokers and smokers' induced sputum [10,28,50].

SP-A, SP-B and SP-D SNPs have been shown to associate with COPD, highlighting a potential causative role in COPD pathogenesis [51-54]. We, for the first time, report lower SP-B concentrations in BAL from ex-smoking COPD subjects *vs.* matched healthy controls, adding to a prior small study of decreased BAL SP-B in a mixture of COPD current and ex-smokers *vs.* healthy controls [55]. SP-B has key roles in surfactant packaging, recycling and maintaining surfactant structure function and SP-B deficiency is lethal [12]. Lower SP-B concentrations could therefore have important physiological consequences.

We demonstrate lower BAL SP-A concentrations in COPD, building on two prior contradictory studies in lung tissue from subjects with moderate COPD which report higher or lower SP-A expression *vs.* healthy controls [27,56]. We also confirm lower BAL SP-D concentrations in COPD [24]. SP-A and SP-D have key roles in lung homeostasis through neutralising, opsonising, agglutinating and clearing pathogens, particles and apoptotic cells, whilst preventing aberrant inflammatory pathways and damage to the delicate lung epithelium[19]. Increased utility of these roles in COPD could potentially deplete levels of these proteins [19,51,57,58]. SP-D knock-out mice develop an emphysematous phenotype with impaired surfactant regulation, influx of inflammatory cells and increased apoptotic cells, metalloproteinases and cytokines, highlighting that low SP-D levels could contribute to emphysema pathogenesis. SP-A and SP-D knock-out models demonstrate increased susceptibility to an array of respiratory viruses and bacteria with associated host-mediated inflammation following infection or allergen challenge[19], highlighting the potential importance of our finding of deficient SP-A and SP-D in COPD on risk of infectious exacerbations and inflammation. Delivery of a recombinant SP-D fragment largely resolves the emphysematous phenotype in the SP-D knock-out mice, as well as the susceptibility to respiratory pathogens, raising the potential of this as a novel therapeutic [19,59,60].

Surfactant has been proposed to be dysregulated in respiratory diseases through various mechanisms [19]. Damage to AT2 cells through noxious stimuli and AT2 cell loss through emphysematous changes and alteration to lung parenchymal architecture could lead to decreased surfactant production [23]. Our data supports this by showing an association between decreased surfactant and emphysema. Due to the nature of bronchoscopy sampling within our study, it was not possible to directly sample the distal airways to look at AT2 gene expression or to correlate transcriptomics of surfactant genes with our findings.

Future studies with paired samples taken from resected lung tissue will be important and could add further clarity to the relative contribution of gene expression *vs.* other mechanisms involved in our finding of surfactant regulation in COPD.

Inflammation-related damage to the delicate epithelial-endothelial barrier could also lead to surfactant loss through leakage into the blood. [24,61-63] This aligns with our findings of higher serum SP-D in COPD, as well as the negative correlation between serum and BAL SP-D levels. Serum SP-B was observed in some donors, predominantly COPD subjects and two HV-ES with hiatus hernia, and was absent in HV-NS. Despite its more hydrophilic nature, SP-A was not detected in serum, potentially due to its larger size. Increased serum SP-B and SP-D in COPD, complemented by corresponding BAL findings of reduced surfactant proteins and lipids, suggest surfactant proteins could have utility as lung-specific peripheral biomarkers for COPD. We investigated this through our logistic regression analysis on serum SP-D levels, which had good predictive value for COPD status and demonstrated the potential of SP-D in particular as a COPD biomarker.

We found neutrophil elastase to be increased in COPD *vs.* control BAL, and inversely correlated with SP-A and SP-D concentrations. Neutrophil elastase, alongside other host and pathogen-associated enzymes, degrades SP-A and SP-D and imbalances could contribute to decreased SP-A and SP-D levels in the COPD lung [19,46,47,64-66]. Potential altered lipid metabolism and surfactant catabolism by alveolar macrophages in COPD could also lead to altered surfactant turnover and dysregulation.[19,51,57,58]. In contrast to a recent study of the COPD alveolar macrophage transcriptome [48], we did not see signs of altered lipid metabolism in alveolar macrophage expression signatures. However, this previous study included more severe and predominantly smoking COPD patients, which may explain

differences with our observations. We demonstrated higher levels of ceramides in COPD *vs.* HV-NS. Ceramides have been reported to influence surfactant production and activity and could therefore lead to surfactant dysregulation [67]. However, we did not see differences in other sphingolipids, as have previously been reported [68,69].

We recognise that associations may not indicate causation and that it is impossible to fully rule out other potentially confounding clinical parameters or pathological mechanisms. We normalised our BAL analyses for protein content and performed statistical testing to address potential confounding effects. However, there are additional factors, which may be difficult to completely address. Inhaled therapeutics have been reported to influence surfactant regulation[70] , and surfactant proteins SP-A, SP-B and SP-D expression have been reported to be increased by corticosteroids[71]. In this study, we see lower levels of these proteins in COPD subjects, a large proportion of whom were on inhaled therapies. Furthermore, although, samples were frozen and stored at -80^{o}C and were analysed immediately after thawing, we cannot rule out the potential for endogenous enzymes to have degraded components and influenced our results. Due to intensive study sampling, our well characterised cohort was mild and relatively small, making it impossible to rule out false negative findings in macrophage lipid metabolism. SP-C was below the limit of detection in both BAL and serum. However, SP-C has previously been reported to be the human lung surfactant protein of lowest abundance by weight [72,73]. Due to cohort heterogeneity and the distinct nature of the various omics datasets, we used an adjusted p-value for the BAL multiomic analysis and trends in surfactant-associated protein. However, lipid downregulation was significantly different between COPD and healthy ex-smokers and never-smokers. Notably the proteome coverage reported here, may vary from previous reports due to differences in methodology, analysis stringency, cohort composition, volume instilled into the lungs and sample volume

[74,75]. In our study, we perform TMT mass spectrometry analysis on peptides originating from 25 µl of BAL and detect >900 proteins per donor sample. In contrast, Tu *et al*. perform BAL proteome profiling across twenty donors, using 10 ml of BAL each and quantify 423 proteins, less than half the proteome coverage we achieved from 400x more sample input. Cohort size is reflective of donor heterogeneity and while studies with small populations may identify more differentially abundant proteins, they have limited statistical power and disease representation. Furthermore, here we report proteomic findings from TMT and label-free mass spectrometry analysis of matched BAL and serum, respectively, that provided a global, unbiased view of proteome dynamics in COPD that may be overlooked by targeted, antibody-based applications despite potentially higher protein identification rates. Our comprehensive multiomics study used a deeply phenotyped and well-characterised mild-moderate COPD cohort to demonstrate global dysregulation of surfactant in the COPD lung, which was associated with emphysematous changes and airway obstruction. Longitudinal studies in early disease and different COPD endotypes will add further clarity to the causes of surfactant dysregulation, the impact on disease progression and importantly the potential for novel surfactant-replacement and surfactant-targeting therapeutics for the future.

Acknowledgements

For MICA II Study group see Online Supplement. The authors thank all the study volunteers for their contribution towards furthering knowledge about chronic obstructive pulmonary disease. They also thank the nursing staff in the Southampton Centre for Biomedical Research. The authors thank VIDA for the image analysis which formed part of an academic collaboration. The authors thank Grith L. Sørensen and Uffe Holmskov (University of

Southern Denmark) for provision of the biotinylated mouse anti-human SP-D (Hyb246-04) antibody.

Statements

Availability of data and materials

The datasets generated and analysed during the current study are not publicly available in order to protect the privacy of all individuals whose data we have collected, stored, and analysed. However, data may be made available upon reasonable request by applying through the established Data Request Portal through which Researchers can request access to de-identified clinical data (https://vivli.org), after which, clinical data may be made available upon review of the patient consent forms, scientific merit of the proposal, and signature of a data sharing/collaboration agreement. This mechanism allows controlled, risk-managed accessibility of the data and at the same time safeguards subjects' confidentiality. Proteomic and lipidomic data will be made available as required by the journal.

Conflicts of interest Disclosure

This project was funded by AstraZeneca. Ventzislava A. Hristova, Raghothama Chaerkady, Matthew S. Glover, Bastian Angermann, Graham Belfield, Maria G. Belvisi, Damla Etal, Sonja Hess, Michael Hühn, Christopher McCrae, Daniel Muthas, Steven Novick, Kristoffer Ostridge, Lisa Öberg, Adam Platt, Junmin Wang, report being employees of AstraZeneca and holding AstraZeneca employee stocks and/or stock options. Alex Mackay reports being an employee of AstraZeneca during the conduct of the study and an employee of Novartis upon submission of this article. Novartis played no role and made no contribution, financial or otherwise, to the work in this manuscript. Outi Vaarala reports being an employee of AstraZeneca from 1.8.2014-3.6.2019 and an employee of OrionPharma from June 10, 2019; during the conduct of this study; and owning AstraZeneca stock. Karl J. Staples reports

receiving grants from AstraZeneca within the submitted work. Tom M.A Wilkinson reports grants and personal fees from AstraZeneca, during the conduct of the study; personal fees and other from MMH, grants and personal fees from GSK, personal fees from BI, grants and personal fees from Synairgen, outside the submitted work. Alastair Watson, Jodie Ackland, Hannah Burke, Doriana Cellura, Howard Clark, Anna Freeman, Emily Hall, Ashley I Heinson, Jens Madsen, Anthony D. Postle and C. Mirella Spalluto report having no conflicts of interest.

Figure Legends

Figure 1. BAL Lipidomic analysis showed reduced levels of PC phospholipids, specifically PC 32:0 (DPPC) as well as PG in COPD. **A.** Volcano plots of lipid abundance in HV-NS *vs.* HV-ES (left), HV-ES *vs.* COPD subjects (middle) and HV-NS *vs.* COPD subjects (right). The x-axis displays log-2fold changes, and the y-axis displays –log10 adjusted p-values. The dashed horizontal line represents an adjusted p-value threshold of 0.05. Dipalmitoylphosphatidylcholine (DPPC) is labeled. Lipid classes, including PC (phosphatidylcholines), PG (phosphatidylgycerols), PI (phosphatidylinositols), PE (phosphatidylethanolamines) and TG (triglycerides) are colored. **B.** Covariate-adjusted box plots showing the summed abundance of PC, PG and PI compared across COPD and HV-ES/HV-NS cohorts. **C.** Covariate-adjusted box plot showing the composition of the top three most abundant PC lipids. **D.** Covariate-adjusted box plots showing the composition of the top three most abundant PG lipids. In **B**, **C** and **D**, HV-ES and HV-NS represent ex-smokers and never-smokers, respectively. The significance levels are given by the number of asterisks (** adj. p-value<0.01, * adj. p-value<0.05, NS = not significant). For details regarding covariate adjustment see supplementary methods.

Figure 2. Correlation analysis showed correlation between BAL PC, PG, PI, SFTPA, SFTPB, SFTPD and NAPSA and FEV1/FVC in COPD. SFTPA, SFTPB, SFTPD, CTSH, NAPSA, CD44, ELANE, and MMP9 encode SP-A, SP-B, SP-D, cathepsin H, napsin A aspartic peptidase, neutrophil elastase, and matrix metallopeptidase 9, respectively. The color of each voxel of the heatmap represents the calculated Spearman's correlation coefficient between a COPD lung

function parameter and the abundance of a surfactant protein, the abundance of a surfactant-associated protein or the summed abundance of a lipid category. The y-axis displays protein symbols or lipid abbreviations, and the x-axis displays lung function parameters. FEV1, FVC, FEV1/FVC, %LAA and TLCO% stand for forced expiratory volume, forced vital capacity, the ratio of forced expiratory volume to forced vital capacity, the percentage ratio of low attenuation area to corresponding lung area and the transfer factor for carbon monoxide, respectively. The significance levels are given by the number of asterisks (**adj. p-value<0.01, *adj. p-value<0.05).

Figure 3. BAL Proteomic analysis showed lower surfactant proteins and proteins involved in surfactant synthesis and secretion in COPD. **A.** Volcano plots of protein abundance in HV-NS *vs.* HV-ES (left), HV-ES *vs.* COPD subjects (middle) and HV-NS *vs.* COPD subjects (right). The x-axis displays log2-fold changes, and the y-axis displays -log10 adjusted p-values. The dashed horizontal line represents an adjusted p-value threshold of 0.05. Proteins whose abundance is significantly altered in COPD compared to HV-ES and HV-NS donors (-log10 adjusted p-value >1.3) are labeled on the volcano plots. These include surfactant and surfactant-associated proteins SFTPA, SFTPB, SFTPD, CTSH, NAPSA, CD44, ELANE, and MMP9, which encode SP-A, SP-B, SP-D, cathepsin H, napsin A aspartic peptidase, CD44 antigen, neutrophil elastase, and matrix metallopeptidase 9, respectively. **B.** Covariate-adjusted box plots showing the abundance of surfactant proteins SFTPA, SFTPB, SFTPD and NAPSA, CTSH, CD44, ELANE and MMP9 across COPD and HV-ES/HV-NS cohorts. ES and NS represent ex-smokers and never-smokers, respectively. The significance levels are given by the number of asterisks (**adj. p-value<0.01, *adj. p-value<0.05). For details regarding covariate adjustment see

supplementary methods. **C.** ELISA showing concentrations of BAL SP-D. ELISA was performed using a rabbit polyclonal anti-recombinant fragment of human SP-D (anti-rfhSP-D) antibody capture antibody and biotinylated mouse anti-human SP-D antibody detection antibody with streptavidin-horse radish peroxidase. Quantification was through comparison with a recombinant full length human SP-D standard.

Figure 4. Serum proteomic analysis detected increased SFTPD and SFTPB, which encode SP-D and SP-B, respectively in COPD patients. **A**. Box plot of serum SFTPD abundance in HV-NS, HV-ES and COPD subjects. ES and NS represent ex-smokers and never-smokers, respectively. The significance levels are given by the number of asterisks (**p-value<0.01, *p-value<0.05). The y-axis is the intensity corresponding to SFTPD for each donor where it was detected. **B**. Spearman's rank correlation of serum and donor-matched BAL SFTPD abundance across the cohort. **C**. Two-by-two contingency table for missing values in SFTPB in relation to disease status (i.e., COPD, HV-ES, HV-NS). P-values are obtained from Fisher's exact test, which compares the proportion of missing values in COPD and ES/NS cohorts. The left table compares COPD and ES cohorts, whereas the right table compares COPD and NS cohorts. **D**. Spearman's rank correlation of serum and donor-matched BAL CTSH abundance across cohort. **E**. ROC-curve of the logistic regression model trained on all donor-matched serum samples.

Tables

Table 1

	Control			**COPD**	
	HV-NS	HV-ES	P-Value (HV-NS *vs*. HV-ES)		P-Value (HV-ES controls *vs.* COPD)
N of subjects (Total=62)	16	20	-	26	-
M/F	9/7	11/9	0.9402	20/6	0.1159
Age	63.5 (9.5)	67.5 (6.75)	0.0871	70.0 (9.75)	>0.9999
Pack-years of smoking	0.1 (1.6)	25.0 (18.6)	**<0.0001**	40.5 (37.8)	0.3009
BMI, kg/m2	27.6 (4.6)	27.7 (3.6)	>0.9999	28.5 (5.4)	>0.9999
FEV1%	104.5 (13.5)	100.5 (11.75)	>0.9999	74.0 (14.75)	**<0.0001**
FEV1/FVC ratio	• 79.5 (5.0)	77.5 (4.5)	0.8400	58.5 (15.0)	**<0.0001**
TLCO%	95.5 (15.5)	89.5 (9.25)	0.9127	73.0 (23.0)	**0.0196**
HRCT LAA%	5.32 (4.165)	5.86 (4.98)	0.8877	11.96 (8.68)	**0.0062**
HRCT E/I MLD	0.800 (0.048)	0.800 (0.060)	>0.9999	0.870 (0.080)	**0.0034**
N (%) in ICS	0 (0)	0 (0)	-	14 (53.65)	**8.33E-05**
N (%) on bronchodilators	0 (0)	1 (5.00)	0.3681	20 (76.92)	**1.21E-06**

Table 1. Demographics of healthy volunteer ex-smoker controls compared with COPD.

Data are presented as median and IQR (interquartile range) unless otherwise indicated. Statistical testing performed using Chi-squared test for categorical variables (Sex; Male/Female) and Kruskal-Wallis with Dunn's post hoc for continuous variables (all other variables) This table is similar to other research previously reported in the MICAII population [29-32]

[a] *Definition of abbreviations: BMI = body mass index; COPD = chronic obstructive pulmonary disease, FEV 1 = forced expiratory volume in one second, FVC = forced vital capacity, HV-ES = healthy volunteer never-smoker, HV-NS = healthy volunteer ex-smoker, TLCO% = percent of predicted transfer factor for*

carbon monoxide, %LAA = High-resolution computed tomography determined emphysema measured by % low attenuation areas (%LAA). ICS = inhaled corticosteroids.

	Control			**COPD**	
	HV-NS	HV-ES	P-Value (HV-NS *vs*. HV-ES)		P-Value (HV-ES controls *vs.* COPD)
% Macrophages	32.93 (22.10)	36.42 (14.84)	0.4538	32.48 (19.63)	0.3003
% Neutrophils	2.25 (3.45)	1.20 (2.51)	>0.9999	0.88 (3.89)	>0.9999
% Eosinophils	0.48 (0.44)	0.60 (1.15)	>0.9999	0.22 (0.25)	0.3492
% Lymphocytes	0.08 (1.20)	1.02 (2.03)	0.0975	0.03 (0.85)	**0.0340**
% Epithelial Cells	63.95 (16.33)	57.4 (10.25)	0.3202	62.18 (15.31)	0.2307
% Squamous cells	0.16 (0.51)	0.33 (0.60)	>0.9999	0.23 (0.69)	>0.9999

Table 2. BAL cell counts.

Showing average percentage macrophages, neutrophil, eosinophils, lymphocytes, epithelial cells and squamous cells of total BAL cell counts. Data are presented as median and interquartile rant (IQR). Statistical testing performed using Kruskal-Wallis with Dunn's post hoc.

References

1. WHO. The global burden of disease: 2004 update. 2004. www.who.int/healthinfo/global_burden_disease/2004 report_update/en/. Accessed on May 17, 2021.
2. Burke H, Wilkinson TMA. Unravelling the mechanisms driving multimorbidity in COPD to develop holistic approaches to patient-centred care. *European Respiratory Review.* 2021;30(160):210041.
3. Kong CW, Wilkinson TMA. Predicting and preventing hospital readmission for exacerbations of COPD. *ERJ Open Research.* 2020;6(2):00325-02019.
4. Watson A, Wilkinson TMA. Digital healthcare in COPD management: a narrative review on the advantages, pitfalls, and need for further research. *Therapeutic Advances in Respiratory Disease.* 2022;16:17534666221075493.
5. Canadas O, Olmeda B, Alonso A, Perez-Gil J. Lipid-Protein and Protein-Protein Interactions in the Pulmonary Surfactant System and Their Role in Lung Homeostasis. *Int J Mol Sci.* 2020;21(10).
6. Agassandian M, Mallampalli RK. Surfactant phospholipid metabolism. *Biochim Biophys Acta.* 2013;1831(3):612-625.
7. Batenburg JJ. Surfactant phospholipids: synthesis and storage. *Am J Physiol.* 1992;262(4 Pt 1):L367-385.
8. Holm BA, Wang Z, Egan EA, Notter RH. Content of dipalmitoyl phosphatidylcholine in lung surfactant: ramifications for surface activity. *Pediatr Res.* 1996;39(5):805-811.
9. Haagsman HP, van Golde LM. Synthesis and assembly of lung surfactant. *Annu Rev Physiol.* 1991;53:441-464.
10. Agudelo CW, Samaha G, Garcia-Arcos I. Alveolar lipids in pulmonary disease. A review. *Lipids Health Dis.* 2020;19(1):122.
11. Wright JR. Immunoregulatory functions of surfactant proteins. *Nat Rev Immunol.* 2005;5(1):58-68.
12. Dunbar AE, Wert SE, Ikegami M, et al. Prolonged Survival in Hereditary Surfactant Protein B (SP-B) Deficiency Associated with a Novel Splicing Mutation. *Pediatric Research.* 2000;48(3):275-282.
13. Possmayer F, Nag K, Rodriguez K, Qanbar R, Schurch S. Surface activity in vitro: role of surfactant proteins. *Comp Biochem Physiol A Mol Integr Physiol.* 2001;129(1):209-220.
14. Perez-Gil J. Structure of pulmonary surfactant membranes and films: the role of proteins and lipid-protein interactions. *Biochim Biophys Acta.* 2008;1778(7-8):1676-1695.
15. Conkright JJ, Apsley KS, Martin EP, et al. Nedd4-2-mediated ubiquitination facilitates processing of surfactant protein-C. *Am J Respir Cell Mol Biol.* 2010;42(2):181-189.
16. Fakih D, Pilecki B, Schlosser A, et al. Protective effects of surfactant protein D treatment in 1, 3-β-glucan-modulated allergic inflammation. *American Journal of Physiology-Lung Cellular and Molecular Physiology.* 2015;309(11):L1333-L1343.
17. Littlejohn JR, da Silva RF, Neale WA, et al. Structural definition of hSP-D recognition of Salmonella enterica LPS inner core oligosaccharides reveals alternative binding modes for the same LPS. *PLoS One.* 2018;13(6):e0199175.

18. Ujma S, Carse S, Chetty A, et al. Surfactant protein a impairs genital HPV16 pseudovirus infection by innate immune cell activation in a murine model. *Pathogens.* 2019;8(4):288.
19. Watson A, Madsen J, Clark H. SP-A and SP-D: dual functioning immune molecules with antiviral and 1 immunomodulatory properties. *Frontiers in Immunology.* 2020.
20. Watson A, Phipps MJ, Clark HW, Skylaris C-K, Madsen J. Surfactant proteins A and D: trimerized innate immunity proteins with an affinity for viral fusion proteins. *Journal of innate immunity.* 2019;11(1):13-28.
21. Nayak A, Dodagatta-Marri E, Tsolaki A, Kishore U. An Insight into the Diverse Roles of Surfactant Proteins, SP-A and SP-D in Innate and Adaptive Immunity. *Frontiers in Immunology.* 2012;3(131).
22. Watson A, Kronqvist N, Spalluto CM, et al. Novel expression of a functional trimeric fragment of human SP-A with efficacy in neutralisation of RSV. *Immunobiology.* 2017;222(2):111-118.
23. Ingenito EP, Tsai LW, Majumdar A, Suki B. On the Role of Surface Tension in the Pathophysiology of Emphysema. *American Journal of Respiratory and Critical Care Medicine.* 2005;171(4):300-304.
24. Winkler C, Atochina-Vasserman EN, Holz O, et al. Comprehensive characterisation of pulmonary and serum surfactant protein D in COPD. *Respir Res.* 2011;12:29.
25. Ohlmeier S, Vuolanto M, Toljamo T, et al. Proteomics of human lung tissue identifies surfactant protein A as a marker of chronic obstructive pulmonary disease. *J Proteome Res.* 2008;7(12):5125-5132.
26. Baekvad-Hansen M, Nordestgaard BG, Dahl M. Surfactant protein B polymorphisms, pulmonary function and COPD in 10,231 individuals. *Eur Respir J.* 2011;37(4):791-799.
27. Liu Z, Chen S, Xu Y, Liu X, Xiong P, Fu Y. Surfactant protein A expression and distribution in human lung samples from smokers with or without chronic obstructive pulmonary disease in China. *Medicine (Baltimore).* 2020;99(7):e19118.
28. Agudelo CW, Kumley BK, Area-Gomez E, et al. Decreased surfactant lipids correlate with lung function in chronic obstructive pulmonary disease (COPD). *PLOS ONE.* 2020;15(2):e0228279.
29. Day K, Ostridge K, Conway J, et al. Interrelationships among small airways dysfunction, neutrophilic inflammation, and exacerbation frequency in COPD. *Chest.* 2021;159(4):1391-1399.
30. Watson A, Spalluto CM, McCrae C, et al. Dynamics of IFN-β responses during respiratory viral infection. Insights for therapeutic strategies. *American journal of respiratory and critical care medicine.* 2020;201(1):83-94.
31. Ostridge K, Gove K, Paas KHW, et al. Using Novel Computed Tomography Analysis to Describe the Contribution and Distribution of Emphysema and Small Airways Disease in Chronic Obstructive Pulmonary Disease. *Annals of the American Thoracic Society.* 2019;16(8):990-997.
32. Watson A, Öberg L, Angermann B, et al. Dysregulation of COVID-19 related gene expression in the COPD lung. *Respiratory research.* 2021;22(1):1-13.
33. Ostridge K, Gove K, Paas KHW, et al. Using Novel Computed Tomography Analysis to Describe the Contribution and Distribution of Emphysema and Small Airways Disease in Chronic Obstructive Pulmonary Disease. *Ann Am Thorac Soc.* 2019;16(8):990-997.
34. Burke H, Freeman A, Cellura DC, et al. Inflammatory phenotyping predicts clinical outcome in COVID-19. *Respiratory Research.* 2020;21(1):245.

35. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods.* 2016;13(9):731-740.
36. Mao S, Chaerkady R, Yu W, et al. Resistance to Pyrrolobenzodiazepine Dimers Is Associated with SLFN11 Downregulation and Can Be Reversed through Inhibition of ATR. *Mol Cancer Ther.* 2021;20(3):541-552.
37. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J Lipid Res.* 2008;49(5):1137-1146.
38. Tsugawa H, Ikeda K, Takahashi M, et al. A lipidome atlas in MS-DIAL 4. *Nat Biotechnol.* 2020;38(10):1159-1163.
39. Leth-Larsen R, Nordenbaek C, Tornoe I, et al. Surfactant protein D (SP-D) serum levels in patients with community-acquired pneumonia. *Clin Immunol.* 2003;108(1):29-37.
40. Leth-Larsen R, Garred P, Jensenius H, et al. A common polymorphism in the SFTPD gene influences assembly, function, and concentration of surfactant protein D. *J Immunol.* 2005;174(3):1532-1538.
41. Clements JA. Surface tension of lung extracts. *Proc Soc Exp Biol Med.* 1957;95(1):170-172.
42. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics.* 2008;9(1):1-13.
43. Xiang X, Yuan F, Zhao J, et al. Deficiency in pulmonary surfactant proteins in mice with fatty acid binding protein 4-Cre-mediated knockout of the tuberous sclerosis complex 1 gene. *Exp Physiol.* 2013;98(3):830-841.
44. Dong Y, Arif AA, Guo J, et al. CD44 Loss Disrupts Lung Lipid Surfactant Homeostasis and Exacerbates Oxidized Lipid-Induced Lung Inflammation. *Frontiers in Immunology.* 2020;11(29).
45. Bratcher PE, Weathington NM, Nick HJ, Jackson PL, Snelgrove RJ, Gaggar A. MMP-9 cleaves SP-D and abrogates its innate immune functions in vitro. *PLoS One.* 2012;7(7):e41881.
46. Rubio F, Cooley J, Accurso FJ, Remold-O'Donnell E. Linkage of neutrophil serine proteases and decreased surfactant protein-A (SP-A) levels in inflammatory lung disease. *Thorax.* 2004;59(4):318-323.
47. Hirche TO, Crouch EC, Espinola M, et al. Neutrophil Serine Proteinases Inactivate Surfactant Protein D by Cleaving within a Conserved Subregion of the Carbohydrate Recognition Domain *. *Journal of Biological Chemistry.* 2004;279(26):27688-27698.
48. Fujii W, Kapellos TS, Bassler K, et al. Alveolar macrophage transcriptomic profiling in COPD shows major lipid metabolism changes. *ERJ Open Res.* 2021;7(3).
49. Otto-Verberne C, Ten Have-Opbroek A, Willems L, Franken C, Kramps J, Dijkman J. Lack of type II cells and emphysema in human lungs. *European Respiratory Journal.* 1991;4(3):316-323.
50. Telenga ED, Hoffmann RF, Ruben tK, et al. Untargeted lipidomic analysis in chronic obstructive pulmonary disease. Uncovering sphingolipids. *Am J Respir Crit Care Med.* 2014;190(2):155-164.
51. Sorensen GL. Surfactant Protein D in Respiratory and Non-Respiratory Diseases. *Frontiers in Medicine.* 2018;5(18).
52. Silveyra P, Floros J. Genetic variant associations of human SP-A and SP-D with acute and chronic lung injury. *Front Biosci (Landmark Ed).* 2012;17:407-429.

53. Guo X, Lin H-M, Lin Z, et al. Surfactant protein gene A, B, and D marker alleles in chronic obstructive pulmonary disease of a Mexican population. *European Respiratory Journal.* 2001;18(3):482-490.
54. Guan J, Liu X, Xie J, et al. Surfactant protein a polymorphism is associated with susceptibility to chronic obstructive pulmonary disease in Chinese Uighur population. *J Huazhong Univ Sci Technolog Med Sci.* 2012;32(2):186-189.
55. Um SJ, Lam S, Coxson H, Man SF, Sin DD. Budesonide/formoterol enhances the expression of pro Surfactant Protein-B in lungs of COPD patients. *PLoS One.* 2013;8(12):e83881.
56. Ohlmeier S, Vuolanto M, Toljamo T, et al. Proteomics of human lung tissue identifies surfactant protein A as a marker of chronic obstructive pulmonary disease. *J Proteome Res.* 2008;7(12):5125-5132.
57. Wright JR. Immunoregulatory functions of surfactant proteins. *Nature Reviews Immunology.* 2005;5(1):58-68.
58. Dong Q, Wright JR. Degradation of surfactant protein D by alveolar macrophages. *American Journal of Physiology-Lung Cellular and Molecular Physiology.* 1998;274(1):L97-L105.
59. Knudsen L, Ochs M, MacKay R, et al. Truncated recombinant human SP-D attenuates emphysema and type II cell changes in SP-D deficient mice. *Respiratory Research.* 2007;8(1):70.
60. Watson A, Sørensen GL, Holmskov U, Whitwell HJ, Madsen J, Clark H. Generation of novel trimeric fragments of human SP-A and SP-D after recombinant soluble expression in E. coli. *Immunobiology.* 2020;225(4):151953.
61. Robin M, Dong P, Hermans C, Bernard A, Bersten AD, Doyle IR. Serum levels of CC16, SP-A and SP-B reflect tobacco-smoke exposure in asymptomatic subjects. *Eur Respir J.* 2002;20(5):1152-1161.
62. Leung JM, Mayo J, Tan W, et al. Plasma pro-surfactant protein B and lung function decline in smokers. *Eur Respir J.* 2015;45(4):1037-1045.
63. Papaioannou AI, Konstantelou E, Papaporfyriou A, et al. Serum Surfactant Protein Levels in Patients Admitted to the Hospital with Acute COPD Exacerbation. *Lung.* 2018;196(2):201-205.
64. Ostridge K, Williams N, Kim V, et al. Relationship between pulmonary matrix metalloproteinases and quantitative CT markers of small airways disease and emphysema in COPD. *Thorax.* 2016;71(2):126-132.
65. Ostridge K, Williams N, Kim V, et al. Distinct emphysema subtypes defined by quantitative CT analysis are associated with specific pulmonary matrix metalloproteinases. *Respiratory Research.* 2016;17(1):92.
66. Southampton UH. ACCORD Trial Platform. 2020; https://www.accord-trial.org/. Accessed 29.06.21, 2021.
67. Tibboel J, Reiss I, de Jongste JC, Post M. Ceramides: a potential therapeutic target in pulmonary emphysema. *Respiratory Research.* 2013;14(1):96.
68. Koike K, Berdyshev EV, Mikosz AM, et al. Role of Glucosylceramide in Lung Endothelial Cell Fate and Emphysema. *Am J Respir Crit Care Med.* 2019;200(9):1113-1125.
69. Berdyshev EV, Serban KA, Schweitzer KS, Bronova IA, Mikosz A, Petrache I. Ceramide and sphingosine-1 phosphate in COPD lungs. *Thorax.* 2021;76(8):821-825.

70. Sims MW, Tal-Singer RM, Kierstein S, et al. Chronic obstructive pulmonary disease and inhaled steroids alter surfactant protein D (SP-D) levels: a cross-sectional study. *Respir Res.* 2008;9:13.
71. Tillis CC, Huang HW, Bi W, Pan S, Bruce SR, Alcorn JL. Glucocorticoid regulation of human pulmonary surfactant protein-B (SP-B) mRNA stability is independent of activated glucocorticoid receptor. *Am J Physiol Lung Cell Mol Physiol.* 2011;300(6):L940-950.
72. Weaver TE, Whitsett JA. Function and regulation of expression of pulmonary surfactant-associated proteins. *Biochemical Journal.* 1991;273(Pt 2):249.
73. Hickling TP, Clark H, Malhotra R, Sim RB. Collectins and their role in lung immunity. *J Leukoc Biol.* 2004;75(1):27-33.
74. Tu C, Mammen MJ, Li J, et al. Large-scale, ion-current-based proteomics investigation of bronchoalveolar lavage fluid in chronic obstructive pulmonary disease patients. *J Proteome Res.* 2014;13(2):627-639.
75. Pastor MD, Nogal A, Molina-Pinelo S, et al. Identification of oxidative stress related proteins as biomarkers for lung cancer and chronic obstructive pulmonary disease in bronchoalveolar lavage. *Int J Mol Sci.* 2013;14(2):3440-3455.

MICAII Study Group

MICAII Study Group Members

Ventzislava A. Hristova[1*],

Alastair Watson[2,3*],

Raghothama Chaerkady[1],

Matthew S. Glover[1],

Jodie Ackland[2],

Bastian Angerman[4],

Graham Belfield[5],

Maria G. Belvisi[6,7],

Hannah Burke[2,8],

Doriana Cellura[2],

Damla Etal[5],

Anna Freeman[2,8],

Ashley I Heinson[2],

Sonja Hess[1],

Michael Hühn[4],

Alex Mackay[4,7],

Christopher McCrae[6],

Daniel Muthas[6],

Steven Novick[5],

Kristoffer Ostridge[2,4],

Lisa Öberg[4],

Adam Platt[11],

Anthony D. Postle[2],

C. Mirella Spalluto[2],

Outi Vaarala[12],

Junmin Wang[1],

Karl J. Staples[2,8],

Tom M.A Wilkinson[2,8 #],

Stephanie Ashenden[5],

Sarah Bawden[8],

Aurelie Bornot[5],

Jerome Bouquet[13],

Carolina Caceres[13],

Chia-Chien Chiang[14],

Kerry Day[2,8],

Antonio DiGiandomenico[13],

Hanna Duàn[6],

Vancheswaran Gopalakrishnan[13],

Alex Hicks[2,8],

Fredrik Karlsson[5],

Shameer Khader[14],

Glenda Lassi[6],

Christopher Morehouse[13],

Karl Nordström[5],

Esther Nyimbili[8],

Laura Presland[8],

Nicola Rayner[8],

Pedro Rodrigues[8],

Bret Sellman[13],

Gary Sims[6],

Andria Staniford[8],

Paul Warrener[13],

Nicholas P. Williams[2,8],

Wen Yu[14],

Xiaotao Qu[14],

Bairu Zhang[5],

Tianhui Zhang[5],

Natalie van Zuydam[5],

Bruce Thompson[15],

Ulrika Edvardsson[16],

Stephen Harden[17],

Affiliations for MICAII Study Group

[1]Dynamic Omics, Centre for Genomics Research (CGR), Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, USA

[2]Faculty of Medicine, University of Southampton, Southampton, UK

[3]School of Clinical Medicine, University of Cambridge, Cambridge, UK

[4]Translational Science and Experimental Medicine, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[5]Translational Genomics, Discovery Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[6]Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[7]NHLI, Imperial College, London, UK

[8]NIHR Southampton Biomedical Research Centre, University Hospital, Southampton, UK

[9]Translational Science and Experimental Medicine, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, USA

[10]Data Sciences & Quantitative Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, USA

[11]Translational Science and Experimental Medicine, Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

[12]Faculty of Medicine, University of Helsinki, Helsinki, Finland

[13]Microbial Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[14]Data Science and Artificial Intelligence, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[15]Swinburne University of Technology Melbourne, Australia

[16]Business Development and Licensing, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

[17]University Hospital Southampton NHS Foundation Trust, Southampton, UK

#Nominated consortia representative:

Professor Tom Wilkinson

Faculty of Medicine, University of Southampton, Southampton, UK

t.wilkinson@soton.ac.uk

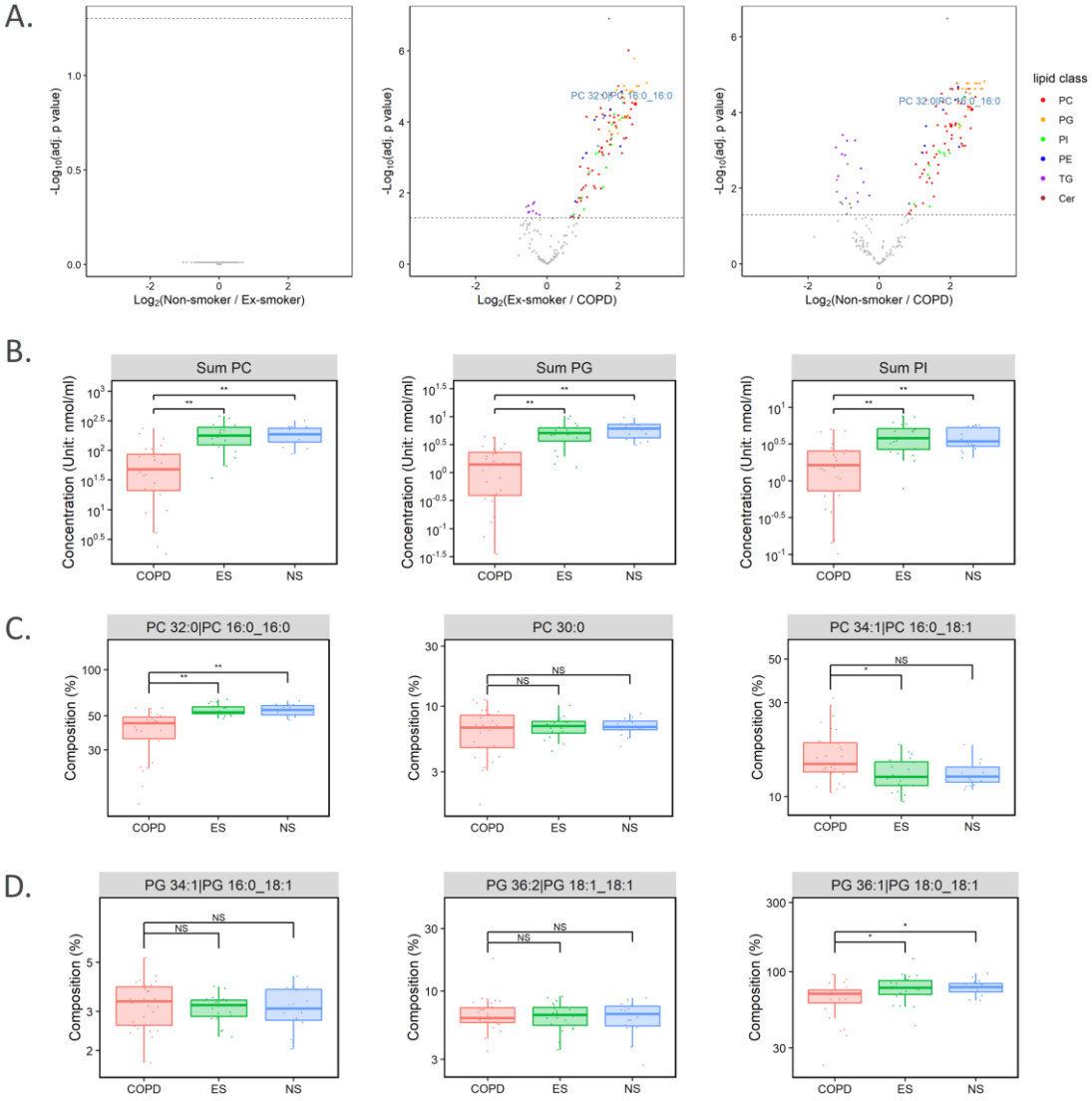

A.
-Log10(adj. p value)
Log2(Non-smoker / Ex-smoker)
Log2(Ex-smoker / COPD)
Log2(Non-smoker / COPD)
PC 32:0|PC 16:0_16:0
lipid class
PC
PG
PI
PE
TG
Cer
B.
Sum PC
Sum PG
Sum PI
Concentration (Unit: nmol/ml)
COPD
ES
NS
C.
PC 32:0|PC 16:0_16:0
PC 30:0
PC 34:1|PC 16:0_18:1
Composition (%)
D.
PG 34:1|PG 16:0_18:1
PG 36:2|PG 18:1_18:1
PG 36:1|PG 18:0_18:1

Figure 1

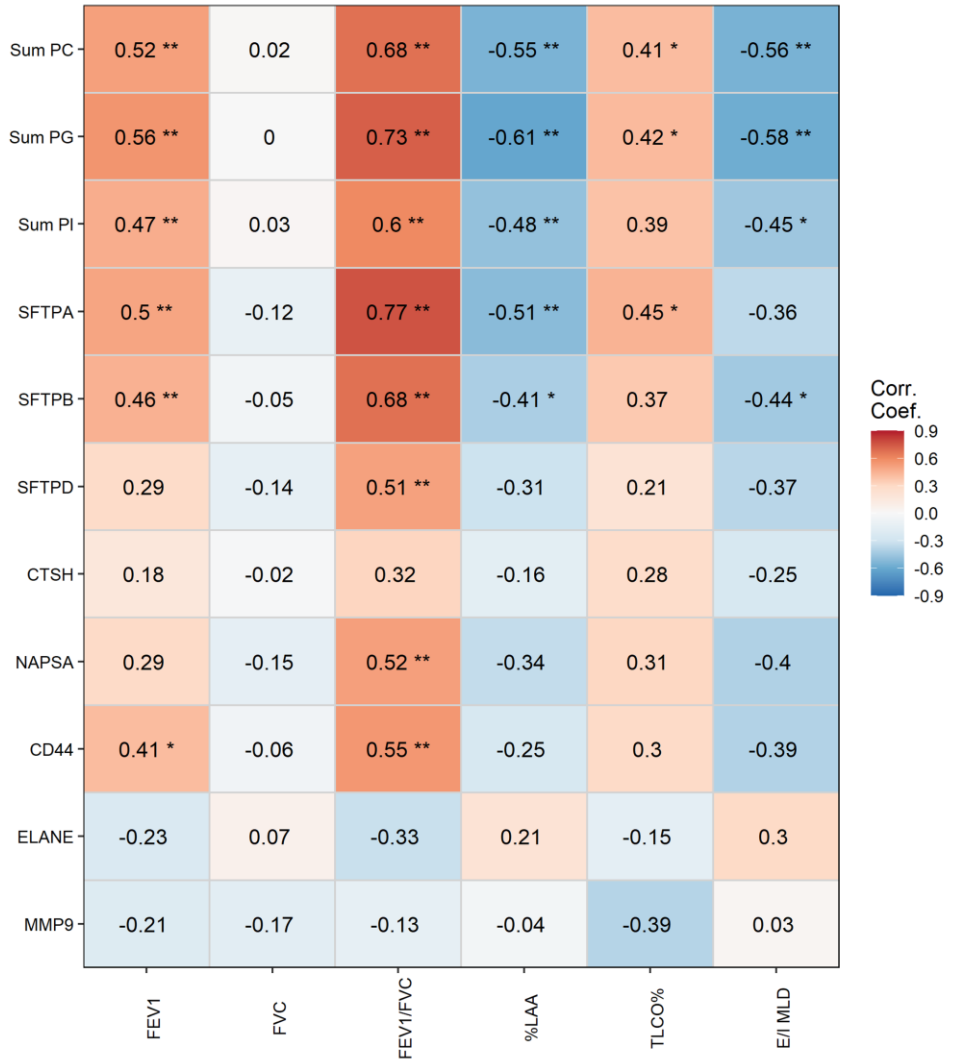

Sum PC
Sum PG
Sum PI
SFTPA
SFTPB
SFTPD
CTSH
NAPSA
CD44
ELANE
MMP9
0.52 ** 0.02 0.68 ** -0.55 ** 0.41 * -0.56 **
0.56 ** 0 0.73 ** -0.61 ** 0.42 * -0.58 **
0.47 ** 0.03 0.6 ** -0.48 ** 0.39 -0.45 *
0.5 ** -0.12 0.77 ** -0.51 ** 0.45 * -0.36
0.46 ** -0.05 0.68 ** -0.41 * 0.37 -0.44 *
0.29 -0.14 0.51 ** -0.31 0.21 -0.37
0.18 -0.02 0.32 -0.16 0.28 -0.25
0.29 -0.15 0.52 ** -0.34 0.31 -0.4
0.41 * -0.06 0.55 ** -0.25 0.3 -0.39
-0.23 0.07 -0.33 0.21 -0.15 0.3
-0.21 -0.17 -0.13 -0.04 -0.39 0.03
FEV1
FVC
FEV1/FVC
%LAA
TLCO%
E/I MLD
Corr. Coef.
0.9
0.6
0.3
0.0
-0.3
-0.6
-0.9

Figure 2

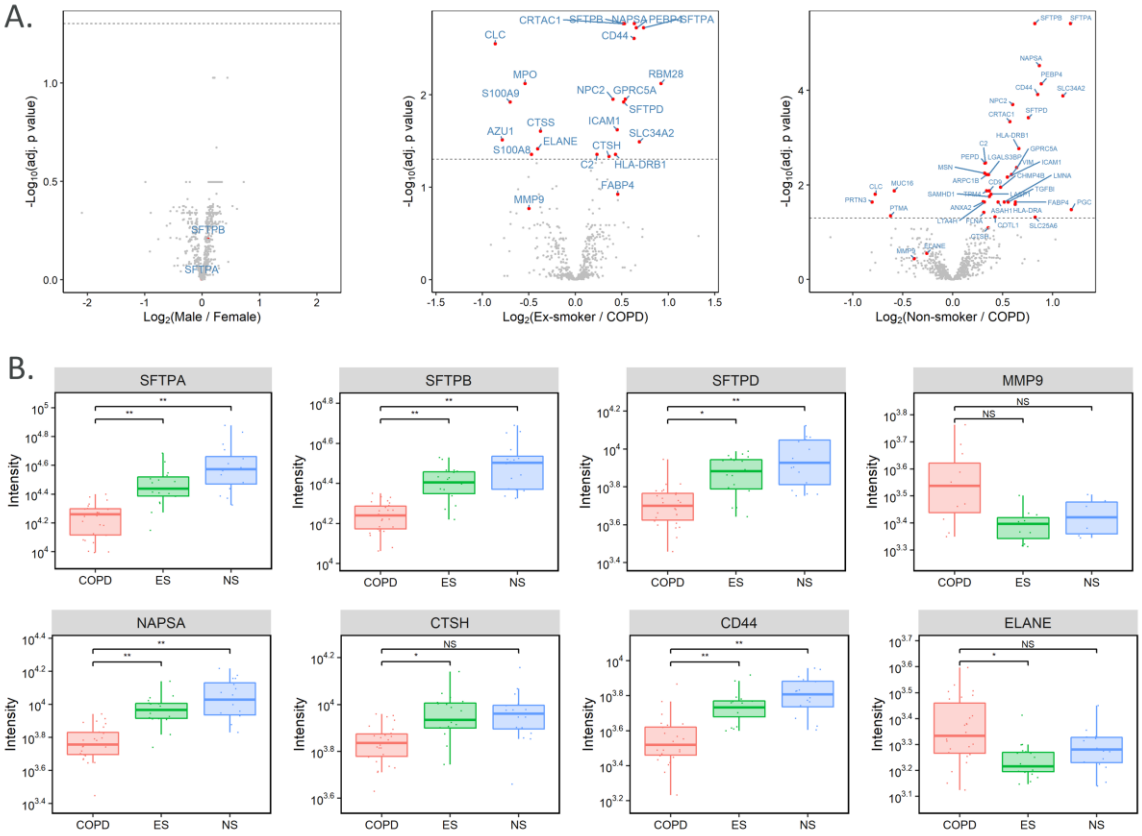
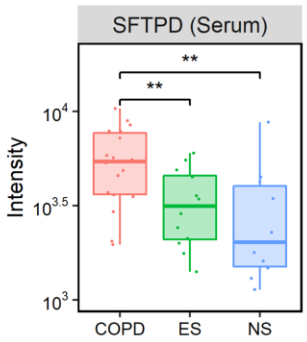
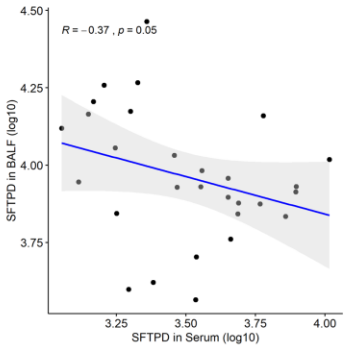

A.
-Log10(adj. p value)
Log2(Male / Female)
SFTPB
SFTPA
Log2(Ex-smoker / COPD)
CRTAC1
SFTPB
NAPSA
PEBP4
SFTPA
CD44
CLC
MPO
RBM28
S100A9
NPC2
GPRC5A
SFTPD
ICAM1
AZU1
CTSS
SLC34A2
ELANE
S100A8
CTSH
C2
HLA-DRB1
FABP4
MMP9
Log2(Non-smoker / COPD)
B.
SFTPA
SFTPB
SFTPD
MMP9
NAPSA
CTSH
CD44
ELANE
Intensity
COPD
ES
NS
C.
SP-D Concentration in BALF (ng/mL)
**p<0.01
NS
COPD
HES
HNS

Figure 3

A.



B.

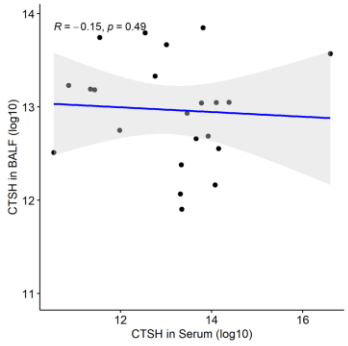


C.

SP-B	COPD	HV-ES	Total
Missing	29	20	49
Not Missing	6	2	8
Total	35	22	
p-value = 0.47			

SP-B	COPD	HV-NS	Total
Missing	29	19	48
Not Missing	6	0	6
Total	35	19	
p-value = 0.08			

D.



E.

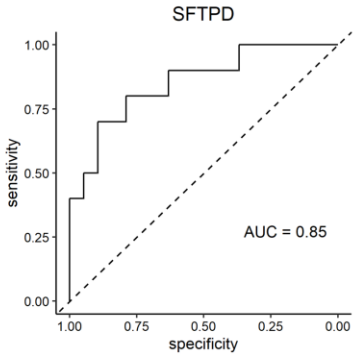


Figure 4

Online Supplement

Supplementary methods

BAL supernatant processing for LC-MS/MS analysis

BAL supernatants for proteomic analysis were processed using an S-Trap-based method (protifi.com). Samples were visually inspected for hemolysis and specimens that appeared pink were omitted from the study. Pilot analysis determined protein content per volume BAL to be consistent across healthy and COPD donors.Fifty microliters of BAL from healthy donors (including never-smokers and ex-smokers) and COPD subjects were treated with S-trap buffer (5% sodium dodecyl sulfate, 50 mM triethylammonium bicarbonate (TEABC) buffer, 0.76% phosphoric acid, pH 7.55) and sonicated for 15 min to completely denature proteins. Subsequently samples were reduced in 10 mM tris(2-caroxyethyl)phosphine hydrochloride (TCEP, Sigma), pH 7.8 for 30 min at 65°C followed by alkylation using 40 mM iodoacetamide (IAA, Sigma) in the dark for 30 min at room temperature. Proteins were digested using sequence grade trypsin/lysC (Promega, Madison, WI) at a 15:1 ratio at 37°C for 12 hrs on micro S-Trap cartridges. The resulting peptides were resuspended in 0.1% trifluoracetic acid (TFA, Sigma), desalted using Oasis HLB 96-well plate (2 mg sorbent, 30µm, Waters) and used for tandem mass tag (TMT) (cat. No. A34808, Thermo Fisher Scientific) labelling according to manufacturer's instructions. Digested peptides derived from 25µl equivalent of 110 BAL supernatants from healthy and COPD donors, were randomized across 11 batches and labelled with 145 µg of 11-plex TMT reagents. Each TMT set contained a similar distribution of healthy non-smoker, ex-smoker and COPD samples and TMT 131C was dedicated to a pooled sample comprised of all study specimens. 11plex-TMT labelled samples were then combined, concentrated in a SpeedVac and fractionated on an Oasis plate (Waters # 186001828BA) under basic conditions. Initially 12 different elutions were collected by using a step gradient of acetonitrile containing 10mM TEABC. Distant fractions were then pooled to generate 3 final samples from each TMT batch for mass spectrometry analysis.

Nanoflow LC-MS/MS analysis

LC-MS/MS analysis of TMT labelled peptides was carried out on a Q Exactive HF-X (Thermo Fisher Scientific) mass spectrometer interfaced with a Dionex 3000 RSLCnano system. Peptides were captured on a 2 cm x 75 µm C18 trap column (ReproSil-Pur 120 C18-AQ 7um) and samples were separated on a monolithic column (50 cm, cut from a 2 m long column, 100 µm ID, GL Sciences Inc. USA) using a gradient of solvent A (0.2% formic acid) and solvent B (0.2% formic acid in 90% acetonitrile). Peptides were separated using a 90 min gradient of solvent B as follows: 4% to 16.5%B in 2.5 - 52.5 min; 33.5% B in 73 min followed by a stay at 98% B for 3 min and re-equilibration at 2% B. A flowrate of 0.7 µL/min was used. Peptides were sprayed in an electrospray ionization (ESI) source using a stainless steel emitter with 2kV at a capillary temperature of 275°C. A full-scan MS spectrum was collected at 60,000 resolution at m/z of 200 and scanned at 350-1200 m/z with automatic gain control (AGC) of 3E6. The top 12 precursors were selected, and an MS/MS scan was obtained at 7,500 resolution with 50 ms injection time, isolation window of 0.9 m/z with offset 0.1 m/z, normalized collision energy (NCE) of 29. For MS2, minimum AGC target was set to 1.7E4. Dynamic exclusion duration was set to 15 sec. The fixed first mass was set to 100 m/z. Charge state exclusion was set to ignore unassigned, 1, and 7 and greater charges. For internal mass calibration, lock mass of 371.10124 m/z was used.

Data analysis

Mass spectrometry data was analysed using Proteome Discoverer 2.3 (Thermo Fisher Scientific) software with search engine Mascot (version 2.6.0). Data was searched using latest UniProt Human protein database. Unfragmented precursor and TMT reporter ions were removed using a non-fragment filter in the PD 2.3 workflow. Search parameters included 3 missed cleavages for trypsin, oxidation (M) and deamidation (N, Q) as variable modifications. Tandem label (229.163Da) at N-terminus and lysine residues and carbamidomethylation on cysteine residues were set as fixed modifications. The mass tolerances on precursor and fragment masses were set at 20 ppm and 0.05 Da, respectively for MS2 analysis. Consensus step in PD2.3 included several nodes for spectrum, peptide and protein grouping and FDR calculation. Reporter ions for TMT labelled peptides were quantified using the PD quantitation node and peak integration tolerance was set at 20 ppm

by considering most confident centroid peaks. Signal to noise values were calculated in addition to measurement of intensities of the TMT reporter ion for peptide and protein quantitation. The intensities were normalized by total peptide amount in PD 2.3. To account for protein input, the global quantitative proteome data was reviewed before normalization and no samples showed an unexpected pattern of distribution. Albumin and hemoglobin abundances were not significantly different between sub-cohorts. Further normalization of the data across all samples was carried out using Reporter Ion Quantitation in Proteome Discoverer, which calculates the total sum of the abundance values for each TMT channel over all peptides identified within a file. The channel with the highest total abundances serves as a reference for correcting abundances across the remaining channels by a constant factor.

Macrophage expression analysis

The RNA-sequencing was conducted as a total RNA-seq using the Kapa RNA HyperPrep Kit with RiboErase, and a paired-end sequencing approach (2 x 51) on an Illumina NovaSeq 6000 platform. Fastq files were processed, quality checked and estimated read counts as well as variance-stabilized transformed data generated. All as been previously been described (1). The average read depth per macrophage sample were 55.9 million. Statistical analysis of the transcriptomic data set was explored using differential expression testing and Weighted Gene Correlation Network analysis (WGCNA) (2). Differential expression testing was performed using DESeq2 (v1.26.0) using apeglm (3) for fold change shrinkage, all in R (v3.6.1). Estimated counts was used as input for DESeq2 with lowly expressed genes excluded (required at least 10 counts in at least 20 samples). In the models used to assess differential expression between subject groups, effects from gender and a technical batch-effect (library batch effect) were taken into account. The Benjamini-Hochberg multiple testing correction method was applied. Weighted Gene Correlation Network Analysis (WGCNA) was also implemented to explore this transcriptomics dataset. WGCNA was performed using the WGCNA R package(2). Variance-stabilized transformed genes expression data were used as input for this analysis. Construction of the gene network was performed using the WGCNA automatic network construction method, which is a 1-step network construction and module detection function. A soft thresholding power of 7 was chosen based on the scale-free topology fit output of the *pickSoftThreshold* function.

Parameters used to cluster genes into modules included *minModuleSize* =50, *mergeCutHeight* = 0.25 and *deepSplit* = 2. Module clustering using module eigengenes was used to identify relationships between modules. Assessment of module association with clinical trait metadata was performed to determine the presence of modules with high trait significance, which may indicate presence of genes or pathways of biological relevance. Gene list enrichment analysis was performed on gene lists extracted from modules or module clusters of interest using ToppFunn, which is part of the online ToppGene Suite using default parameter settings (FDR multiple correction method and enrichment significance cut off level 0.05) (4).

Serum processing for LC-MS/MS proteomic analysis

Serum total protein has minimal inter-individual variability and is highly consistent across donors. The serum proteome has a broad concentration range spanning ~11 orders of magnitude with albumin accounting for more than half the total protein in circulation. In this study, 10 µl of serum per donor underwent depletion of the top fourteen most abundant blood proteins using High Select Top14 Abundant Protein Depletion Kit (Thermo Fisher Scientific) according to manufacturer's instructions. Depleted serum was subjected to reduction, alkylation and trypsin/lysC digestion via EasyPep 96 MS Sample Prep Kit (Thermo Fisher Scientific) as outlined by the manufacturer. All serum samples were processed in a single 96-well EasyPep plate eliminating batch effects. Resultant peptides were dried and resuspended in 0.1% formic acid aqueous buffer.

Serum proteomic nanoflow LC-MS/MS analysis

Serum LC-MS/MS analysis was carried out on an Exploris 480 (Thermo Fisher Scientific) mass spectrometer interfaced with a Dionex 3000 RSLCnano system. Peptides, 150 ng per sample, were injected on an Acclaim PepMap RSLC 75 µm x15 cm column (Thermo Fisher Scientific) at a flow rate of 350 µl/min and separated over a 45 min gradient of solvent A (0.1% formic acid) and solvent B (0.1% formic acid in acetonitrile). Gradient of solvent B as follows: 4% to 24% in 2.5 - 40 min; 36% B at 48 min; 64% B at 48.5 held for 4.5 min; 98% B at 53.5 min held for 1.5 min followed by re-equilibration at 4% B. Peptides were sprayed in an electrospray ionization (ESI) source using a stainless-steel emitter with 1650V at a temperature of 270°C. A full-scan MS spectrum was collected at 120,000 resolution at m/z of 200 and scanned at 350-1200 m/z with automatic gain control (AGC) of 100% corresponding to 1E6. DIA MS/MS scans were obtained at 30,000 resolution with the isolation window set to 21 m/z and an overlap of 1 m/z over a precursor mass range of 350-1200 m/z. AGC target was set to 2000% (100% = 1E5).

Serum proteomic data analysis

Serum DIA analysis was conducted in Spectronaut v15 (Biognosys) using the latest UniProt Human protein database and a spectral library representative of healthy and COPD serum proteomes. Serum DIA raw files were searched against a spectral library generated from data-dependent acquisition (DDA) raw files from five pooled and fractionated serum samples, comprised of ten donors each (five male and five female) merged with DDA serum data from this cohort (non-fractionated). The final spectral library used for this analysis contained 3075 proteins (1417 protein groups) representative of 50 healthy donors (25 male and 25 female) in addition to this entire study cohort. Analysis was performed without imputation, with an FDR=0.01 (Qvalue cut off). All observations that passed the Qvalue threshold at least once were included. A list of protein groups identified in each sample and their corresponding intensities was exported to Perseus for further statistical and graphical analysis.

Lipidomics sample preparation

Lipid extraction from BAL supernatants were performed using a modified Maytash method (5). Frozen BAL aliquots (50 µL) were thawed at 4 °C and mixed for 30 seconds at 2000 RPM and 4 °C. 225 µL cold (-30 °C) methanol (MeOH) was added to samples on ice and mixed for 1 minute at 2000 RPM. Samples were spiked with 1 µL Splash Lipidomix (Avanti Polar Lipids) comprised of 14 deuterated lipid internal standards and mixed for 1 minute at 2000 RPM and 4 °C. 750 µL of methyl tert-butyl ether (MTBE) was added and samples were mixed at 2000 RPM for 6 minutes at 4°C. 187.5 µL H_2O was added to induce phase separation and samples were mix at 2000 RPM for 6 minutes at 4°C. Centrifugation was performed for 5 minutes at 14,000 x*g* and 20 °C. Aliquots of 650 µL lipid supernatant were transferred into separate tubes. Samples were dried in a SpeedVac (Thermo Scientific). Dried extracts were stored at -80 °C until reconstitution and subsequent LC-MS and LC-MS/MS analysis.

LC-MS and LC-MS/MS lipidomics analysis

Lipid fractions were reconstituted in 100 µL 90:10 MeOH:toluene and mixed for 1 minute at 1500 RPM. Samples were sonicated for 2 minutes in a water bath, mixed for 1 minute at 1500 RPM, and centrifuged for 5 minutes at 16,000 x*g* and 20°C. 5 µL from each sample was combined to serve as pooled quality control (QC) sample. Samples were transferred to glass HPLC vials and analysed by LC-MS and LC-MS/MS. Lipid samples were analysed in both

positive and negative mode ionization. Samples were analysed on a Vanquish UHPLC – Orbitrap ID-X Tribrid MS (Thermo Scientific) using a chromatographic method adopted from Fiehn and coworkers (6). Mobile phase A and B were 0.1% formic acid and 10 mM ammonium formate in 60:40 ACN:H_2O and 0.1% formic acid and 10 mM ammonium formate in 90:10 IPA:ACN. Chromatographic separation was performed on Acquity UPLC CSH C18 column (1.7 µm, 2.1 x 100mm) (Waters Corporation). Column temperature was maintained at 65°C. LC-MS analysis was performed with a scan range of 120-1200 m/z at Orbitrap resolution of 60,000 on each individual sample. Lipid identification was performed by LC-MS/MS using HCD fragmentation with AcquireX DeepScan data-dependent acquisition workflow (ThermoFisher) performed on iterative injections of a pooled lipid extract from this study.

Lipidomics data analysis

Lipidomics LC-MS and LC-MS/MS data were analyzed using MS-DIAL version 4.60 (7). Peak detection, adduct assignment, identification, alignment, and normalization were performed in MS-DIAL. Lipid annotations were performed using LipidBlast *in silico* fragmentation spectral library provided with MS-DIAL version 4.60 with all lipid classes considered. Lipids were annotated from LC-MS/MS data with identification score cutoff of 70% and MS and MS/MS mass tolerances of 0.005 Da and 0.05 Da, respectively. Lipid acyl chain compositions are reported as the sum composition for species in which fragmentation spectra does not meet score threshold to confidently assign individual acyl chain compositions (e.g., PC 30:0). The concentration of each lipid was quantified by normalizing to the abundance of SPLASH Lipidomix (Avanti Polar Lipids) isotopically labelled internal standard spiked into each sample (described above) for each lipid class and expressed in nmol/ml. Percent composition of individual lipid species is determined by the ratio of individual lipid species concentrations to the sum of all species identified from the same lipid class (e.g, PCs).

References

1. Watson A, Öberg L, Angermann B, Spalluto CM, Hühn M, Burke H, et al. Dysregulation of COVID-19 related gene expression in the COPD lung. Respiratory research. 2021;22(1):1-13.
2. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics. 2008;9(1):1-13.

3. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35(12):2084-92.
4. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic acids research. 2009;37(suppl_2):W305-W11.
5. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. J Lipid Res. 2008;49(5):1137-46.
6. Cajka T, Smilowitz JT, Fiehn O. Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography-High-Resolution Mass Spectrometry Platforms. Anal Chem. 2017;89(22):12360-8.
7. Tsugawa H, Ikeda K, Takahashi M, Satoh A, Mori Y, Uchino H, et al. A lipidome atlas in MS-DIAL 4. Nat Biotechnol. 2020;38(10):1159-63.

Supplementary Tables

	Control			**COPD**	
	HV-NS	HV-ES	P-Value (HV-NS *vs*. HV-ES)		P-Value (HV-ES controls *vs.* COPD)
N of subjects (Total=62)	19	22	-	34	-
M/F	11/8	12/10	0.8294	26/8	0.0863
Age	63.0 (12.0)	66.5 (7.3)	0.0989	70.0 (11.5)	0.5679
Pack-years of smoking	0.0 (1.6)	25.0 (18.1)	**<0.0001**	40.5 (32.6)	0.1272
BMI, kg/m2	28.0 (5.2)	27.7 (4.2)	>0.9999	28.3 (6.6)	>0.9999
FEV1%	102.0 (15.5)	100.0 (10.75)	>0.9999	78.0 (25.0)	**<0.0001**
FEV1/FVC ratio	80.0 (5.0)	87.5 (4.3)	0.5336	55.0 (17.0)*	**<0.0001**
TLCO%	95.5 (15.5)[&]	89.0 (12.5)[&]	0.4520	73.0 (23.8)[&]	**0.0285**
HRCT LAA%	5.32 (4.17)[^]	5.86 (4.98)[^]	0.6348	13.16 (8.74)[^]	**0.0017**
HRCT E/I MLD	0.800 (0.048)[^]	0.800 (0.060)[^]	>0.9999	0.875 (0.075)[^]	**0.0018**
N (%) in ICS	0 (0)	0 (0)	-	14 (44.18)[$]	**0.00237**
N (%) on bronchodilators	0 (0)	1 (5.00)	**8.33-E05**	20 (70.59)[$]	**1.26-E06**

Table S1. Demographics of cohort included for proteomic analysis of serum

Data are presented as median and IQR (interquartile range) unless otherwise indicated. Statistical testing performed using Chi-squared test for categorical variables (Sex; Male/Female, ICS use or not and bronchodilator use or not) and Kruskal-Wallis with Dunn's post hoc for continuous variables (all other variables) This table is similar to other research previously reported in the MICAII population [26-29]

[a] *Definition of abbreviations: BMI = body mass index; COPD = chronic obstructive pulmonary disease, FEV 1 = forced expiratory volume in one second, FVC = forced vital capacity, HV-ES = healthy volunteer never-smoker, HV-NS = healthy volunteer ex-smoker, TLCO% = percent of predicted transfer factor for carbon monoxide, %LAA = High-resolution computed tomography determined emphysema measured by % low attenuation areas (%LAA). ICS = inhaled corticosteroids.*

*Notably analysis was undertaken on serum samples from subjects from the final MICAII cohort (table 1), in addition to subjects who were removed from the study prior to bronchoscopy due to numerous reasons, including subject request, not being suitable for bronchoscopy, or not fitting the inclusion criteria as set out in the methodology. Some of these subjects therefore did not undergo *lung function assessment (1 COPD), %TLCO assessment (3 HV-NS; 1 HV-ES; 11 COPD), ^HRCT scan (3 HV-NS; 2 HV-ES; 8 COPD) or $inhaled medications were not recorded (1 COPD). These data for these patients are therefore not included within this table.*

Table S2. Serum proteome summary (table attached at end of document due to size)

UniProt ID and corresponding gene name for serum proteins identified across all donors in this cohort.

	Control			**COPD**	
	HV-NS	HV-ES	P-Value (HV-NS *vs*. HV-ES)		P-Value (HV-ES controls *vs.* COPD)
N of subjects (Total=62)	15	18	-	14	-
M/F	9/6	9/9	0.5659	11/3	0.0977
Age	64.0 (7.0)	67.5 (6.80)	0.1457	72.5 (10.5)	0.4193
Pack-years of smoking	0.2 (1.8)	25.0 (20.9)	**<0.0001**	45.0 (40.8)	0.5947
BMI, kg/m2	28.0 (5.2)	28.2 (4.0)	0.9882	29.3 (5.3)	>0.9999
FEV1%	103.0 (17.0)	100.5 (8.8)	0.8526	79.5 (12.3)	**0.0002**
FEV1/FVC ratio	79.0 (4.0)	77.0 (4.8)	0.3541	61.0 (11.3)	**<0.0001**
TLCO%	95.0 (16.0)	88.0 (10.0)	0.3331	81.0 (16.0)	0.3331
HRCT LAA%	5.69 (3.99)	5.38 (4.32)	>0.9999	10.8 (8.03)	**0.0362**
HRCT E/I MLD	0.800 (0.045)	0.795 (0.050)	0.8184	0.840 (0.070)	**0.0112**
N (%) in ICS	0 (0)	0 (0)	-	7 (50.00)	**0.00237**
N (%) on bronchodilators	0 (0)	1 (5.00)	**8.33-E05**	12 (85.71)	**1.26-E06**

Table S3. Demographics of cohort included for transcriptomic analysis of purified BAL macrophages

Data are presented as median and IQR (interquartile range) unless otherwise indicated. Statistical testing performed using Chi-squared test for categorical variables (Sex; Male/Female, ICS use or not and bronchodilator use or not) and Kruskal-Wallis with Dunn's post hoc for continuous variables (all other variables) This table is similar to other research previously reported in the MICAII population [26-29]

[a] *Definition of abbreviations: BMI = body mass index; COPD = chronic obstructive pulmonary disease, FEV 1 = forced expiratory volume in one second, FVC = forced vital capacity, HV-ES = healthy volunteer never-smoker, HV-NS = healthy volunteer ex-smoker, TLCO% = percent of predicted transfer factor for carbon monoxide, %LAA = High-resolution computed tomography determined emphysema measured by % low attenuation areas (%LAA). ICS = inhaled corticosteroids.*

Supplementary figure legends

Figure S1. Gender differences were not significant across omics datasets.
(A) Lipid distribution between male and female donors was not significantly different. **(B)** Proteome profiles of male and female donors were not significantly different, all had a -log10 adjusted p-value <1.3. SFTPA and SFTPB showed minimal differential expression and abundance differences between genders, SFTPA was slightly more abundant in females and SFTPB was slightly more abundant in males.

Figure S2. Correlation analysis of SFTPB, SFTPA and SFTPD with NAPSA, CTSH and neutrophil elastase in BAL.
Spearman's rank correlation of (**A**) SFTPB correlation with napsin A aspartic peptidase (NAPSA), (**B**) SFTPB correlation with Cathepsin H (CTSH), (C) SFTPA correlation with neutrophil elastase (ELANE), (D) SFTPD correlation with ELANE.

Figure S3. Construction of macrophage gene network and detection of modules.
Construction of dendrogram was performed using an automatic one step network construction and module detection method. **(A)** A soft thresholding power of 7 was chosen based on scale-free topology fit indicating the lowest power which intersected the high value red line (R^2 = 0.9) on the scale independence plot, whilst maintaining a mean connectivity score above 0. **(B)** Clustering dendrogram of genes, with dissimilarity based on topological overlap and colours below indicating module assignment. Performed using the *WGCNA* package in R and the additional parameters: *minModuleSize* =50, *mergeCutHeight* = 0.25 and *deepSplit* = 2.

Table S2. Serum proteome summary

UniProt ID	Genes
A0A024R6I7	SERPINA1
A0A075B6I9; P04211	IGLV7-46;IGLV7-43
A0A075B6K0; P01717; P01718	IGLV3-16;IGLV3-25;IGLV3-27
A0A075B7C5; A0A494C1Q1; P13501	CCL5
A0A087WTK0; A0A087WVC6; Q12913; Q12913-2	PTPRJ
A0A087WTY6; A3KFI1; A3KFI2; A3KFI3; A3KFI4; A3KFI5; E5RFZ1; P41271; P41271-2	NBL1
A0A087WV50; A0A087WYT4; A0A0B4J215; C9J2P9; H3BTT7; J3KNA0; S4R3N7	SENP3;STK4;NIN;CBLL1;HERPUD1;OXA1L;C10orf90
A0A087WWU8; P06753-2; P06753-3; P06753-6	TPM3
A0A087WX77; P13591	NCAM1
A0A087WY68; A0A087WZR0; H0Y3Q0; P29122; P29122-2; P29122-7; P29122-8	PCSK6
A0A087WYI3; P41439	FOLR3
A0A087WYS1	UGP2
A0A087WZM2; D6REQ6; D6RHI9; H0YAE9; O00584	RNASET2;RNASET2;RNASET2;;RNASET2
A0A087WZR4; A0A3B3ISU3; H0Y4U3; M9MML6; O75015	FCGR3B
A0A087X054; A0A494C039; Q9Y4L1	HYOU1

A0A087X0D5; P09668	CTSH
A0A087X0M8	CHL1
A0A087X0Q4	IGKV2-40
A0A087X0S5; P12109	COL6A1
A0A087X0T8; A0A087X1W8; A0A0A0MTJ8; Q9BY67; Q9BY67-2; Q9BY67-3; Q9BY67-4; Q9BY67-5; X5DQS5	CADM1
A0A087X1J7; P22352	GPX3
A0A096LPE2	SAA2-SAA4
A0A0A0MRJ7; P12259	F5
A0A0A0MRN7; Q6YP21; Q6YP21-2	KYAT3
A0A0A0MRZ8; P04433	IGKV3D-11;IGKV3-11
A0A0A0MS08; P01857	IGHG1
A0A0A0MS09; P01880; P01880-2	IGHD
A0A0A0MS15	IGHV3-49
A0A0A0MT69	IGKJ4
A0A0A0MTH3; Q13418; Q13418-2; Q13418-3	ILK
A0A0B4J1R4; P32754; P32754-2	HPD
A0A0B4J1R6; P29401; P29401-2	TKT
A0A0B4J1U7	IGHV6-1
A0A0B4J1X5	IGHV3-74
A0A0B4J231; B9A064	IGLL5
A0A0C4DFP6; Q9NQ79; Q9NQ79-2	CRTAC1
A0A0C4DG49; P15151; P15151-2; P15151-3; P15151-4	PVR

A0A0C4DH25	IGKV3D-20
A0A0C4DH34	IGHV4-28
A0A0C4DH67	IGKV1-8
A0A0D9SEN1; Q12884	FAP
A0A0G2JMB2	IGHA2
A0A0G2JMC9; A0A0G2JMW8; A8MZH0; Q8N149; Q8N149-2; Q8N149-3; Q8N149-4	LILRA2
A0A0G2JMW3; A0A0G2JP44; Q9HBB8; Q9HBB8-2; Q9HBB8-4	CDHR5
A0A0G2JMY9; Q8N6C8; Q8N6C8-3	LILRA3
A0A0J9YX35	IGHV3-64D
A0A0J9YXX1	IGHV5-10-1
A0A0J9YY99	
A0A0M3HER1; P48059; P48059-2; P48059-3; P48059-4; P48059-5; Q7Z4I7; Q7Z4I7-2; Q7Z4I7-3; Q7Z4I7-4; Q7Z4I7-5	LIMS1;LIMS1;LIMS1; LIMS1;LIMS1;LIMS1; LIMS2;LIMS2;LIMS2; LIMS2;LIMS2
A0A0S2Z4L3; A0A3B3ISJ1; P07225	PROS1
A0A0U1RQQ4; Q9UNN8	PROCR
A0A140T8Y3; A0A140T902; A0A140TA33; A0A140TA52; A0A3B3ISX9; P22105; P22105-1; P22105-4	TNXB
A0A140TA49	C4A
A0A1B0GV23; A0A1B0GVD5; A0A1B0GWE8; P07339	CTSD

A0A286YES1; A0A4W9A917; P01860	IGHG3
A0A286YEY4; P01859	IGHG2
A0A286YFJ8; P01861	IGHG4
A0A2Q2TTZ9	IGKV1D-33
A0A2R8Y430; P48637	GSS
A0A2R8Y478; A6NNI4; G8JLH6; P21926	CD9
A0A2R8Y524; A0A2R8YEC9; E9PFW2; O00462	MANBA
A0A2R8YEP4; P30043	BLVRB
A0A2U3TZL5; E9PNW4; P13987; P13987-2	CD59;;CD59;CD59
A0A3B3IQ51; P36980; P36980-2	CFHR2
A0A3B3IS66	F13B
A0A3B3IS80; P05062	ALDOB
A0A3B3ISD1; C1KBH7; P11362-19; P11362-21; P11362-7	FGFR1
A0A3B3ISR2; B4DPQ0	C1R
A0A3B3ISS6; Q14956; Q14956-2	GPNMB
A0A3B3ISU0; Q02487; Q02487-2	DSC2
A0A494C0L6; C9JGI3; P19971; P19971-2	TYMP
A0A494C0X7; D3DSM0; P05107	ITGB2
A0A494C165; K7ES25; P12955	PEPD
A0A499FJK2; P01137	TGFB1
A0A4W8ZXM2	IGHV3-72
A0A590UJJ6; B4DEB1; K7EK07; K7EMV3; K7EP01; K7ES00; P68431; P84243; Q16695;	H3-3A;H3-3A;H3-3B;H3-3B;H3-3B;H3-

Q5TEC6; Q6NXT2; Q71DI3	3B;H3C1;H3-3A;HIST3H3;H3-2;H3F3C;HIST2H3A
A0A590UK92; O14746; O14746-2; O14746-3; O14746-4	TERT
A0A5F9UJX7; A0A5F9UP49; G3V1E2; Q9BRK5; Q9BRK5-3; Q9BRK5-4; Q9BRK5-6	SDF4
A0A5F9UY30; A0A5H1ZRP2; O43493; O43493-3; O43493-5; O43493-7	TGOLN2
A0A5F9ZHM4; P07195	LDHB
A0A5H1ZRQ3	IGKC
A0A5H1ZRQ7; A0M8Q6	IGLC7
A1A5D9-2; Q9UJX6-2	BICDL2;ANAPC2
A1L4H1	SSC5D
A6NC48; Q10588	BST1
A6XND0; A6XND1; P17936; P17936-2	IGFBP3
B0QYF7; B0QYF8; F2Z2F1; P02144; Q8WVH6	MB
B0YIW2; P02656	APOC3
B1ALD9; Q15063; Q15063-3; Q15063-5	POSTN
B1AN99; P35030; P35030-2; P35030-3; P35030-4; P35030-5	PRSS3
B1AP13; H3BLV0; H7BY55; P08174; P08174-2; P08174-3; P08174-4; P08174-5; P08174-6; P08174-7	CD55
B1B0D4; Q86TH1	ADAMTSL2

B4DV12; F5GXK7; F5GYU3; F5H265; F5H2Z3; F5H388; F5H6Q2; F5H747; J3QKN0; J3QS39; J3QTR3; P0CG47; P0CG48; P62979; P62987; Q5PY61; Q96C32	UBB;UBC;UBC;UBC;UBC;UBC;UBC;UBC;UBB;UBB;RPS27A;UBB;UBC;RPS27A;UBA52;UBC;UBC
B4E3Q4; Q9NZK5	ADA2
B7Z6Z4; F8W1R7; G3V1V0; G8JLA2; J3KND3; P60660; P60660-2	MYL6
B7ZKJ8	ITIH4
B9A064-2	IGLL5
C9IZP8	C1S
C9J0J0; Q96EE4	CCDC126
C9JF17; P05090	APOD
C9JFR7; P99999	CYCS
C9JL85; P58546	MTPN
C9JZC2	ZNF621
D3YTG3; H0Y897; H7C556; Q7Z7G0; Q7Z7G0-2; Q7Z7G0-3; Q7Z7G0-4	ABI3BP
D6R934; P02746	C1QB
D6RD17; P01591	JCHAIN
D6RE82; P56182	RRP1
D6RE86	CP
D6RF35; P02774; P02774-3	GC
D6RF86; P55285; P55285-2	CDH6
D6RIU5; P00995	SPINK1

D6W5L6; P07988	SFTPB
E5RJD0; H0YBY3; P17900	GM2A
E7END6; P04070; P04070-2	PROC
E7EQB2; E7ER44; P02788; P02788-2	LTF
E7ESB3; Q13508; Q13508-2; Q13508-3	ART3
E7ET86; Q8IVW4; Q8IVW4-2	CDKL3
E7ETH0	CFI
E7EUF1; Q13822; Q13822-2; Q13822-3	ENPP2
E7EUJ1; P11150	LIPC
E7EV71; Q14766; Q14766-2; Q14766-4; Q14766-5	LTBP1
E9PD35; P35916; P35916-1	FLT4
E9PEK4; P07333	CSF1R
E9PEP6; O14786	NRP1
E9PK25; H0Y4A7; P23528	CFL1;BRWD1;CFL1
E9PKY4; Q03167; Q03167-2	TGFBR3
E9PND2; E9PP21; E9PS42; P21291	CSRP1
E9PRU1; H0YET5; O95967	EFEMP2
F5GXJ9; Q13740; Q13740-2	ALCAM
F5GY80; F5H7G1	C8B
F5GZS6; J3KPF3; P08195; P08195-2; P08195-3; P08195-4	SLC3A2
F5GZZ9; Q86VB7; Q86VB7-2; Q86VB7-3	CD163
F5H2B5	PLD4
F5H8B0; P08709; P08709-2	F7

G3V2W1; Q9UK55	SERPINA10
G3V3A0	SERPINA3
G3V4U0; Q9UBX5	FBLN5
G3XAI2; P07942	LAMB1
G3XAK1; P26927	MST1
G5E9Z4	PI4K2B
H0Y2Y8; Q15942; Q15942-2	ZYX
H0Y5E4; H0YCV9; H0YD13; H0YDW7; P16070; P16070-10; P16070-11; P16070-12; P16070-13; P16070-14; P16070-15; P16070-16; P16070-17; P16070-18; P16070-3; P16070-4; P16070-5; P16070-6; P16070-7; P16070-8; P16070-9	CD44
H0Y755; M9MML0; P08637	FCGR3A
H0YAC1; P03952	KLKB1
H0YGX7; P52566	ARHGDIB
H0YJW9	
H0YLC7; P16930; P16930-2	FAH
H3BR24; H3BTD5; M0R1K8	CCPG1;MYZAP;ARHGEF1
H7BY64; Q96NZ9	ZNF511-PRAP1;PRAP1
H7C1K7	KIF15
H7C5R1	CP
H9KV31; O15394	NCAM2
H9N1E7; P07359	FLT1;GP1BA

I3L145; P04278	SHBG
I3L397; I3L504; P63241; P63241-2	EIF5A
J3KNB4; P49913	CAMP
J3KNV4; Q13683; Q13683-10; Q13683-3; Q13683-7; Q13683-9	ITGA7
J3KPA1; P54108; P54108-2; P54108-3	CRISP3
J3QQR8; J3QQX6; J3QRQ1; J3QRT5; P13598	ICAM2
K4DIA0; O75144; O75144-2	ICOSLG
K7ELL7; P14314; P14314-2	PRKCSH
K7ER74; P02655	APOC4-APOC2;APOC2
K7ERG9; P00746	CFD
K7ERI9; P02654	APOC1
M0QY68; Q9BTV5	FSD1
M0QZ43; P23327	HRC
M0R1Q1	C3
M0R3C9; Q9UM47	NOTCH3
O00151	PDLIM1
O00187	MASP2
O00391	QSOX1
O00533	CHL1
O00592; O00592-2	PODXL
O00602	FCN1
O14498	ISLR
O14645-2	DNALI1

O14791; O14791-2	APOL1
O15204; O15204-2	ADAMDEC1
O43157	PLXNB1
O43852; O43852-3	CALU
O43866	CD5L
O60235	TMPRSS11D
O75083	WDR1
O75368	SH3BGRL
O75594	PGLYRP1
O75636	FCN3
O75882-2	ATRN
O76061	STC2
O94985-2	CLSTN1
O95445	APOM
O95479; R4GMU1	H6PD
O95497	VNN1
O95810	CAVIN2
O95980	RECK
P00338; P00338-3	LDHA
P00450	CP
P00488	F13A1
P00491	PNP
P00533; P00533-3; P00533-4	EGFR
P00558; P00558-2	PGK1
P00734	F2

P00738	HP
P00739	HPR
P00740	F9
P00742	F10
P00747	PLG
P00748	F12
P00915	CA1
P00918	CA2
P01008	SERPINC1
P01009	SERPINA1
P01011	SERPINA3
P01019	AGT
P01023	A2M
P01024	C3
P01031	C5
P01033; Q5H9A7	TIMP1
P01034	CST3
P01042	KNG1
P01042-2	KNG1
P01344-3	IGF2
P01624	IGKV3-15
P01700	IGLV1-47
P01714	IGLV3-19
P01766	IGHV3-13
P01780	IGHV3-7

P01833	PIGR
P01834	IGKC
P01871	IGHM
P01876	IGHA1
P02452	COL1A1
P02461; P02461-2	COL3A1
P02647	APOA1
P02649	APOE
P02652; V9GYM3	APOA2
P02671	FGA
P02675	FGB
P02679; P02679-2	FGG
P02730; P02730-2	SLC4A1
P02741	CRP
P02743	APCS
P02745	C1QA
P02747	C1QC
P02748	C9
P02749	APOH
P02750	LRG1
P02751-1; P02751-3	FN1
P02751-10	FN1
P02753; Q5VY30	RBP4
P02760	AMBP
P02763	ORM1

P02765	AHSG
P02766	TTR
P02768	ALB
P02774-2	GC
P02775	PPBP
P02776	PF4
P02787	TF
P02790	HPX
P03950	ANG
P03951	F11
P04003	C4BPA
P04004	VTN
P04066	FUCA1
P04075	ALDOA
P04114	APOB
P04180	LCAT
P04196	HRG
P04217	A1BG
P04275	VWF
P04279; P04279-2	SEMG1
P04406	GAPDH
P04745	AMY1A
P05019; P05019-2; P05019-3; P05019-4	IGF1
P05067; P05067-11; P05067-7; P05067-8; P05067-9	APP

P05106	ITGB3
P05109	S100A8
P05121; P05121-2	SERPINE1
P05154	SERPINA5
P05155; P05155-3	SERPING1
P05160	F13B
P05164; P05164-2; P05164-3	MPO
P05362	ICAM1
P05451	REG1A
P05452	CLEC3B
P05543	SERPINA7
P05546	SERPIND1
P05556	ITGB1
P06276	BCHE
P06312	IGKV4-1
P06331	IGHV4-34
P06396	GSN
P06396-2	GSN
P06681	C2
P06702	S100A9
P06703; R4GN98	S100A6
P06727	APOA4
P06732	CKM
P06899; P23527; P33778; Q16778	H2BC11;HIST1H2BO; HIST1H2BB;HIST2H2

	BE
P07237	P4HB
P07307; P07307-2; P07307-3	ASGR2
P07357	C8A
P07358	C8B
P07360	C8G
P07384	CAPN1
P07437; Q5JP53	TUBB
P07451	CA3
P07737	PFN1
P07858	CTSB
P07911; P07911-4; P07911-5; X6RBG4	UMOD
P07996	THBS1
P07998	RNASE1
P08185	SERPINA6
P08253	MMP2
P08254	MMP3
P08294	SOD3
P08311	CTSG
P08493; P08493-2	MGP
P08514; P08514-2; P08514-3	ITGA2B
P08519	LPA
P08567	PLEK
P08571	CD14

P08581; P08581-2	MET
P08603	CFH
P08697	SERPINF2
P09172	DBH
P09382	LGALS1
P09486	SPARC
P09619	PDGFRB
P09871	C1S
P0C0L4	C4A
P0C0L5	C4B
P0DJI8	SAA1
P0DJI9	SAA2
P0DOY2; P0DOY3	IGLC2;IGLC3
P10124	SRGN
P10451; P10451-2; P10451-3; P10451-4	SPP1
P10586; P10586-2	PTPRF
P10599	TXN
P10643	C7
P10645	CHGA
P10646	TFPI
P10720	PF4V1
P10721; P10721-2	KIT
P10909; P10909-2; P10909-4; P10909-5; P10909-6	CLU
P11021	HSPA5

P11047	LAMC1
P11142	HSPA8
P11226	MBL2
P11279	LAMP1
P11597	CETP
P11717	IGF2R
P12111	COL6A3
P12318; P12318-2	FCGR2A
P12724	RNASE3
P12814	ACTN1
P12830	CDH1
P13473; P13473-2; P13473-3	LAMP2
P13497	BMP1
P13671	C6
P13727	PRG2
P13796	LCP1
P14151; P14151-2	SELL
P14209	CD99
P14543	NID1
P14618-2	PKM
P14625	HSP90B1
P14780	MMP9
P15144	ANPEP
P15169	CPN1
P16035	TIMP2

P16109; Q5R349	SELP
P16284; P16284-2; P16284-3; P16284-4; P16284-5; P16284-6	PECAM1
P17301	ITGA2
P17813; P17813-2	ENG
P17931	LGALS3
P18065	IGFBP2
P18206; P18206-2	VCL
P18428	LBP
P18615-4	NELFE
P19021; P19021-2; P19021-3; P19021-4; P19021-5; P19021-6	PAM
P19022; P19022-2	CDH2
P19320	VCAM1
P19652	ORM2
P19823	ITIH2
P19827	ITIH1
P19827-2	ITIH1
P20023; P20023-2; P20023-3; P20023-4	CR2
P20742	PZP
P20851; P20851-2	C4BPB
P21333; P21333-2	FLNA
P21709	EPHA1
P22692	IGFBP4
P22792	CPN2

P22891; P22891-2	PROZ
P22897	MRC1
P23141; P23141-2; P23141-3	CES1
P23142	FBLN1
P23142-4	FBLN1
P23284	PPIB
P23470; P23470-2	PTPRG
P24592	IGFBP6
P24593	IGFBP5
P24821	TNC
P25311	AZGP1
P25786; P25786-2	PSMA1
P25815	S100P
P26038	MSN
P26992	CNTFR
P27105	STOM
P27169	PON1
P27348	YWHAQ
P27487	DPP4
P27797	CALR
P27918	CFP
P28799; P28799-3	GRN
P29279	CCN2
P29622	SERPINA4
P30101	PDIA3

P30530; P30530-2	AXL
P31146	CORO1A
P31151	S100A7
P31260-2	HOXA10
P32004; P32004-2; P32004-3	L1CAM
P32119	PRDX2
P32942	ICAM3
P33151	CDH5
P34096	RNASE4
P35247	SFTPD
P35443	THBS4
P35555	FBN1
P35579	MYH9
P35590	TIE1
P35858; P35858-2	IGFALS
P36222	CHI3L1
P36955	SERPINF1
P39060; P39060-1; P39060-2	COL18A1
P40197	GP5
P41222	PTGDS
P42785; P42785-2	PRCP
P43121	MCAM
P43251; P43251-2; P43251-3; P43251-4	BTD
P43652	AFM
P46531	NOTCH1

P48723	HSPA13
P48740	MASP1
P48740-2	MASP1
P49747	COMP
P49908	SELENOP
P51884	LUM
P54289; P54289-2; P54289-5	CACNA2D1
P55056	APOC4
P55058	PLTP
P55103	INHBC
P55268	LAMB2
P55290; P55290-4	CDH13
P55774	CCL18
P58335; P58335-2; P58335-3; P58335-4	ANTXR2
P59665; P59666	DEFA1;DEFA3
P61158	ACTR3
P61224; P61224-3	RAP1B
P61626	LYZ
P61769	B2M
P61981	YWHAG
P62328	TMSB4X
P62937	PPIA
P63104	YWHAZ
P67936	TPM4
P68366; P68366-2	TUBA4A

P68871	HBB
P69905	HBA1
P80108	GPLD1
P80188; X6R8F3	LCN2
P80723; P80723-2	BASP1
P80748	IGLV3-21
P81605; P81605-2	DCD
P98160	HSPG2
Q01459	CTBS
Q01518; Q01518-2	CAP1
Q02985	CFHR3
Q03591	CFHR1
Q04756	HGFAC
Q04917	YWHAH
Q06033; Q06033-2	ITIH3
Q06413-6	MEF2C
Q07954	LRP1
Q08380	LGALS3BP
Q10001	
Q12794; Q12794-2; Q12794-3	HYAL1
Q12805; Q12805-2; Q12805-3; Q12805-4	EFEMP1
Q12805-5	EFEMP1
Q12841	FSTL1
Q12860	CNTN1
Q13093	PLA2G7

Q13103	SPP2
Q13201	MMRN1
Q13332; Q13332-2; Q13332-3; Q13332-4; Q13332-6	PTPRS
Q13790	APOF
Q14112; Q14112-2	NID2
Q14116-2	IL18
Q14118	DAG1
Q14126	DSG2
Q14515	SPARCL1
Q14520; Q14520-2	HABP2
Q14624	ITIH4
Q14624-2	ITIH4
Q15067-3; Q3I5F7	ACOX1;ACOT6
Q15113	PCOLCE
Q15166	PON3
Q15293; Q15293-2	RCN1
Q15404; Q15404-2	RSU1
Q15485	FCN2
Q15582	TGFBI
Q15848	ADIPOQ
Q16270; Q16270-2	IGFBP7
Q16610	ECM1
Q16627	CCL14
Q16706	MAN2A1

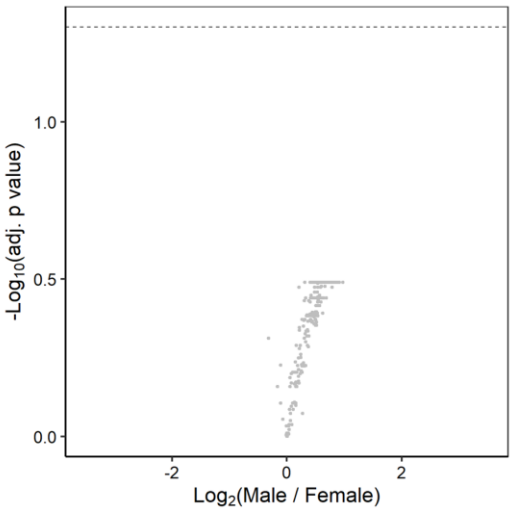
Q16853	AOC3
Q5MJ68	SPDYC
Q5SZC9; Q9P1F3	ABRACL
Q5T123; Q9H299	SH3BGRL3
Q5TFM2	CFH
Q5VY43	PEAR1
Q68G74-2	LHX8
Q6E0U4-16; Q6E0U4-2; Q6E0U4-5; Q6E0U4-6	DMKN
Q6EMK4	VASN
Q6GTS8	PM20D1
Q6UWP8	SBSN
Q6UWP8-2	SBSN
Q6UX71	PLXDC2
Q6UXB8	PI16
Q6UY14; Q6UY14-2; Q6UY14-3	ADAMTSL4
Q6YHK3	CD109
Q71F56	MED13L
Q71U36; Q71U36-2	TUBA1A
Q76LX8	ADAMTS13
Q7Z7M0	MEGF8
Q86U17	SERPINA11
Q86UD1	OAF
Q86UX7; Q86UX7-2	FERMT3
Q86VX2-2	COMMD7
Q86YW5; Q86YW5-2	TREML1

Q86YZ3	HRNR
Q8IWV2	CNTN4
Q8IXL6	FAM20C
Q8IYA8-3	CCDC36
Q8IZF2; Q8IZF2-2	ADGRF5
Q8NBP7	PCSK9
Q8NDA2; Q8NDA2-2; Q8NDA2-4	HMCN2
Q8TDL5	BPIFB1
Q8WWZ8	OIT3
Q8WZ75; Q8WZ75-2	ROBO4
Q92496; Q92496-2	CFHR4
Q92520	FAM3C
Q92743	HTRA1
Q92820	GGH
Q92859; Q92859-2; Q92859-3; Q92859-4	NEO1
Q92954-3	PRG4
Q93063; Q93063-2; Q93063-3	EXT2
Q96IY4	CPB2
Q96KN2	CNDP1
Q96PD5	PGLYRP2
Q99453	PHOX2B
Q99650; Q99650-2	OSMR
Q99784; Q99784-3; Q99784-5	OLFM1
Q99969	RARRES2
Q99972	MYOC

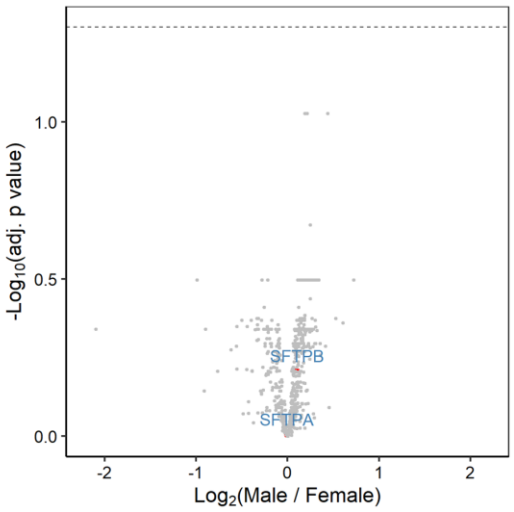
Q9BQ51	PDCD1LG2
Q9BTY2	FUCA2
Q9BUN1	MENT
Q9BWP8; Q9BWP8-10; Q9BWP8-2; Q9BWP8-3; Q9BWP8-4; Q9BWP8-5; Q9BWP8-6; Q9BWP8-7; Q9BWP8-8; Q9BWP8-9	COLEC11
Q9BXR6	CFHR5
Q9BYJ0	FGFBP2
Q9H089; Q9H4S2	LSG1;GSX1
Q9H1U4	MEGF9
Q9H4A9	DPEP2
Q9H4B7	TUBB1
Q9H4G4	GLIPR2
Q9H6X2; Q9H6X2-2; Q9H6X2-4; Q9H6X2-5; Q9H6X2-6	ANTXR1
Q9H8L6	MMRN2
Q9HBR0	SLC38A10
Q9HCB6	SPON1
Q9HDC9	APMAP
Q9NPG4	PCDH12
Q9NPH3; Q9NPH3-2; Q9NPH3-5	IL1RAP
Q9NPR2; Q9NPR2-2	SEMA4B
Q9NPY3	CD93
Q9NQM4	PIH1D3
Q9NY15	STAB1

Q9NZ08; Q9NZ08-2	ERAP1
Q9NZP8	C1RL
Q9NZT1	CALML5
Q9P232	CNTN3
Q9UBG0	MRC2
Q9UBR2	CTSZ
Q9UBX1	CTSF
Q9UEW3; Q9UEW3-2	MARCO
Q9UGM5	FETUB
Q9UHG3	PCYOX1
Q9UIB8; Q9UIB8-2; Q9UIB8-3; Q9UIB8-4; Q9UIB8-5; Q9UIB8-6	CD84
Q9UJC5	SH3BGRL2
Q9UJJ9	GNPTG
Q9UKD1	GMEB2
Q9UKX2	MYH2
Q9ULI3	HEG1
Q9UNW1	MINPP1
Q9Y251; Q9Y251-2	HPSE
Q9Y490	TLN1
Q9Y5C1	ANGPTL3
Q9Y5Y7	LYVE1
Q9Y646	CPQ
Q9Y6R7	FCGBP
Q9Y6Z7	COLEC10

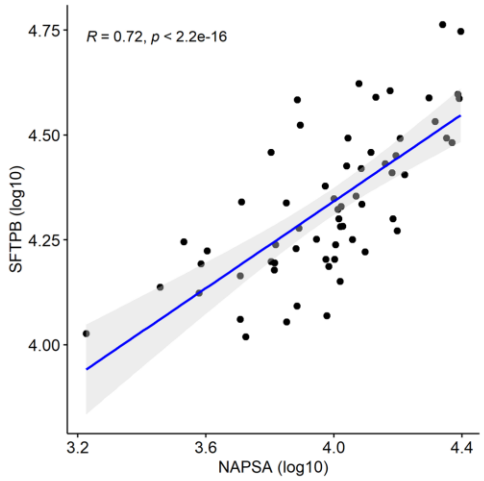
A.


1.0
0.5
0.0
-Log10(adj. p value)
-2
0
2
Log2(Male / Female)

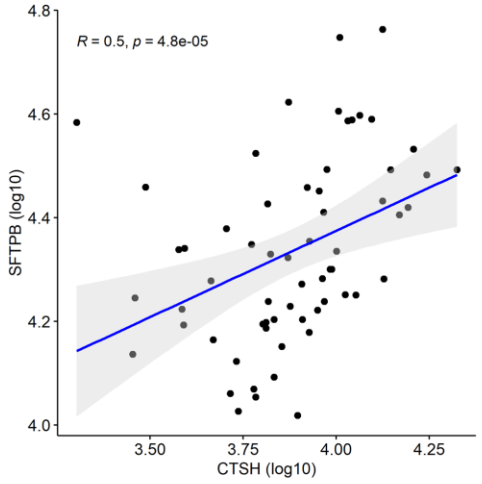
B.


1.0
0.5
0.0
-Log10(adj. p value)
SFTPB
SFTPA
-2
-1
0
1
2
Log2(Male / Female)

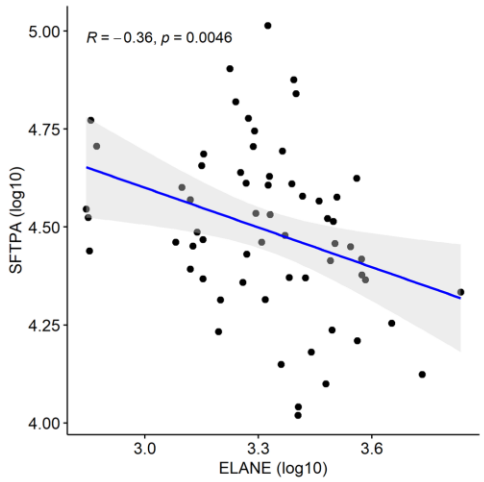
A.


R = 0.72, p < 2.2e-16
SFTPB (log10)
NAPSA (log10)
4.75
4.50
4.25
4.00
3.2
3.6
4.0
4.4

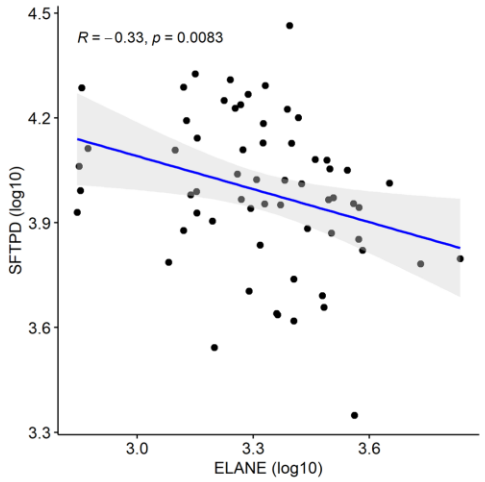
B.

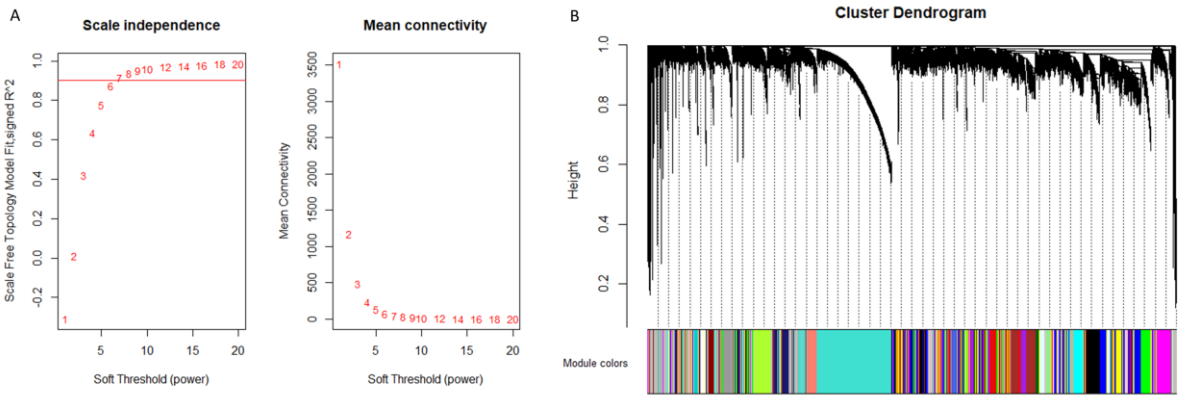

R = 0.5, p = 4.8e-05
SFTPB (log10)
CTSH (log10)
4.8
4.6
4.4
4.2
4.0
3.50
3.75
4.00
4.25

C.


R = −0.36, p = 0.0046
SFTPA (log10)
ELANE (log10)
5.00
4.75
4.50
4.25
4.00
3.0
3.3
3.6

D.


R = −0.33, p = 0.0083
SFTPD (log10)
ELANE (log10)
4.5
4.2
3.9
3.6
3.3
3.0
3.3
3.6


A
Scale independence
Scale Free Topology Model Fit,signed R^2
Soft Threshold (power)
Mean connectivity
Mean Connectivity
Soft Threshold (power)
B
Cluster Dendrogram
Height
Module colors