The lung proteome in HIV-associated obstructive lung disease

Sarah Samorodnitsky[1], Danielle Weise[2], Eric F. Lock[1], Ken M. Kunisaki[2,3], Alison Morris[4], Janice M. Leung[5], Monica Kruk[2], Laurie Parker[6], Pratik Jagtap[6], Timothy J. Griffin[6] and Chris H. Wendt[2,3]

[1]Biostatistics Division, School of Public Health, University of Minnesota, Minneapolis, MN, USA. [2]Department of Medicine, University of Minnesota, Minneapolis, MN, USA. [3]Department of Medicine, Minneapolis VA Health Care System, Minneapolis, MN, USA. [4]Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [5]Department of Medicine, University of British Columbia, Vancouver, Canada. [6]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN, USA.

Corresponding author: Chris Wendt (wendt005@umn.edu)

Shareable abstract (@ERSpublications)
Protein expression differs in the lung of HIV infected individuals with OLD. Pathway analysis reveals cell adhesion molecules and hematopoietic cell lineage cell surface interactions at the vascular wall associating with low lung function. https://bit.ly/4gMU3kj

Cite this article as: Samorodnitsky S, Weise D, Lock EF, *et al*. The lung proteome in HIV-associated obstructive lung disease. *ERJ Open Res* 2025; 11: 00204-2024 [DOI: 10.1183/23120541.00204-2024].

Received: 4 March 2024
Accepted: 13 Sept 2024

Abstract

Rationale Obstructive lung disease is increasingly common among persons living with HIV (PLWH). There are currently no validated biomarkers that identify individuals at risk of developing obstructive lung disease (OLD), and specific mechanisms contributing to HIV-associated OLD remain elusive, independent of smoking. We sought to identify biomarkers and biological pathways associated with OLD using a broad proteomic approach.

Methods We performed tandem mass tagging and mass spectrometry (MS) analysis on bronchoalveolar lavage fluid samples from persons living with HIV with OLD (n=26) and without OLD (n=26). We combined untargeted MS with a targeted SomaScan aptamer-based approach. We used Pearson correlation tests to identify associations between each protein and lung function (forced expiratory volume in 1 s (FEV_1) % pred). We adjusted for multiple comparisons using a false discovery rate adjustment. Significant proteins were entered into a pathway over-representation analysis. Protein-driven endotypes were constructed using K-means clustering.

Measurements and main results We identified over 3800 proteins by MS and identified 254 proteins that correlated with FEV_1 % pred when we combined the MS and SomaScan proteomes when adjusting for smoking status. Pathway analysis revealed cell adhesion molecules as significant.

Conclusions Protein expression differs in the lung of PLWH and decreased lung function (FEV_1 % pred). Pathway analysis reveals cell adhesion molecules having potentially important roles in this process.

Introduction

The use of highly active antiretroviral therapy (HAART) has led to significant declines in morbidity and mortality amongst persons living with HIV (PLWH). However, with the prolongation of life has emerged an increased prevalence of comorbidities, including obstructive lung disease (OLD) [1–8]. Prior to the common use of antiretroviral therapy, OLD was associated with advanced disease (AIDS), previous intravenous drug use and/or pulmonary infections [9]. Subsequently, as pulmonary infections became less dominant, HIV has emerged as an independent contributor to the development of OLD, even independent of smoking. Although proposed mechanisms include immune dysregulation, oxidant stress and dysbiosis [10–13], there are currently no validated biomarkers that identify individuals at risk of developing OLD, and specific mechanisms contributing to HIV-associated OLD remain elusive.

We sought to understand the differences in the proteome between those with and without OLD to identify biomarkers and biological pathways associated with OLD development and worsening lung function. We previously identified proteins and biological pathways expressed in the lung in individuals with HIV-associated OLD using a targeted aptamer-based SomaScan proteomic approach [14]. The goal of this study was to

combine an untargeted mass spectrometry (MS)-based approach with the targeted proteome to provide broader coverage of the lung proteome to identify additional biomarkers and biological pathways of disease.

Methods

Study population

PLWH who had undergone bronchoscopy were selected from the Pittsburgh and Vancouver Lung HIV Cohorts [15, 16]. This consisted of individuals (n=26) with OLD as defined as the ratio of forced expiratory volume in 1 s/forced vital capacity (FEV_1/FVC) <lower limit of normal. Individuals without OLD (n=26) consisted of PLWH with normal lung function (defined as FEV_1/FVC >lower limit of normal and FEV_1 >80% of predicted normal) matched on age (+/− 5 years), antiretroviral treatment use and smoking status (current *versus* nonsmoker). Participants who were in their usual state of health fasted prior to research bronchoalveolar lavage fluid (BALF) sample collection, and pulmonary function tests were performed within 3 months of collecting the samples. Participants in the parent cohort studies provided informed consent for BALF collection and storage with approval by their respective Institutional Review Board at Pittsburgh and Vancouver. The current study was approved by the University of Minnesota Institutional Review Board.

Sample processing

At study enrolment, BALF was collected as previously described [15, 16]. BALF underwent centrifugation at the local collection sites to remove cells, and cell-free BALF samples were stored at −80°C prior to processing for this study. BALF samples from 21 out of 26 with OLD and 24 out of 26 without OLD had adequate protein amounts for tandem mass tagging (TMT, Thermo Fisher Scientific) and MS analysis. The BALF was processed for MS as previously described [17]. Briefly, the cell-free BALF samples were vortexed and centrifuged at 500×G for 10 min at 4°C to separate the soluble BALF components from the insoluble BALF components prior to further analysis. An additional spin of the BALF at 100 000×g was performed to collect a second (hard spin) BALF insoluble component pellet, which was combined with the first BALF pellet. The supernatant, or soluble component of the BALF, was sent for SomaScan analysis as previously reported [14]. The samples, both the soluble component and the insoluble component, were separately processed for TMT labelling and liquid chromatography–mass spectrometry (LC-MS) analysis as previously described [17]. Protein identification of mucin-5AC was performed using parallel reaction monitoring (PRM) (see supplementary methods).

Statistical analysis

The data cleaning process is described in the supplemental methods. We used Global Lung Initiative (GLI)-based race-adjusted calculations of % predicted FEV_1as a measure of lung function capacity. We used Pearson correlation tests and linear regression to identify significant associations between each protein and FEV_1 % pred. Using linear regression, we adjusted for smoking status as a confounder. We refer to these two approaches as the unadjusted and adjusted approaches, respectively. We corrected for multiple comparisons using a false discovery rate (FDR) adjustment [18]. Within the unadjusted and adjusted analyses, we tested for an association between each protein and FEV_1 % pred and applied the FDR adjustment for the soluble component and insoluble component separately. For each dataset, we deemed proteins as significant if the corresponding FDR fell below the 0.1 level. We considered the significant proteins from each dataset separately in a pathway analysis using IMPaLa software with the full set of identified proteins from each dataset as the reference list [19].

We pooled results for the association between proteins and FEV_1 % pred across TMT of the soluble and insoluble components and SomaScan of the soluble component. We did this for the unadjusted and adjusted analyses separately. Taking the unadjusted analysis as an example, we combined p-values resulting from Pearson correlation testing between FEV_1 % pred and proteins recorded across all datasets (soluble component TMT, insoluble component TMT, and SomaScan) using Fisher's method [20]. We then employed a permutation testing framework to evaluate the significance of the association between each protein and FEV_1 % pred. We used a permutation testing framework for two reasons: 1) to ensure robustness against deviations from normality; and 2) to preserve the same dependency structure between the proteins in each dataset and permuted FEV_1 % pred, as these p-values were obtained from overlapping (non-independent) samples. We considered this framework both unadjusted and adjusted for smoking status. To assess the robustness of the detected associations, we also considered diffusing capacity of the lung for carbon monoxide % predicted (D_{LCO} % pred) as an outcome using the same framework. Our permutation testing framework is described in more detail in the supplementary methods.

Lastly, we used two clustering approaches to identify and validate protein-driven endotypes in BALF. We first applied k-means 100 times and used 10 random start points to obtain a robust solution [14]. We applied this approach to the soluble component TMT and insoluble component TMT datasets separately

using all identified proteins in each (see the supplementary methods for details). We also applied Bayesian consensus clustering (BCC), an integrative approach to clustering on multiple datasets [21]. We used BCC to assess the similarities and differences between the clustering schemes on each dataset and the consensus clustering scheme integrated across datasets. We applied BCC to the 44 samples which overlapped in the soluble component TMT, insoluble component TMT and soluble component SomaScan datasets. The datasets included both overlapping and non-overlapping proteins. We allowed each dataset to adhere differently to the overall clustering scheme. We applied BCC using the R package provided in the BayesCC R package (https://github.com/ttriche/bayesCC).

Results

Study participant demographics

Table 1 summarises the demographics of participants in our study. The soluble component TMT and insoluble component TMT datasets differed by one sample with OLD, but overall showed similar demographic distributions across those with and without OLD. Most samples in this cohort were men

TABLE 1 Characteristics of participants included in our study by dataset (TMT BALF soluble component or TMT BALF insoluble component)

	Soluble component TMT samples		Insoluble component TMT samples		Total
	With OLD	Without OLD	With OLD	Without OLD	
Participants n	21	24	21	24	
Sex					
Male	15 (71.4)	17 (70.8)	15 (71.4)	17 (70.8)	33 (71.7)
Female	6 (28.6)	7 (29.2)	7 (28.6)	7 (29.2)	13 (28.3)
Age years[#]					
Mean±SD	58.6±7.72	53.8±7.62	58.9±7.97	53.8±7.62	56.3±8.10
Median (min, max)	58.0 (44.0, 76.0)	54.0 (42.0, 76.0)	58.0 (44.0, 76.0)	54.0 (42.0, 76.0)	56.0 (42.0, 76.0)
Ethnicity					
Black, non-Hispanic	12 (57.1)	12 (50.0)	12 (57.1)	12 (50.0)	25 (54.3)
White, Hispanic/Latino	9 (42.9)	11 (45.8)	9 (42.9)	11 (45.8)	20 (43.5)
Asian/Pacific Islander	0 (0)	1 (4.2)	0 (0)	1 (4.2)	1 (2.2)
Smoking status					
Former	7 (33.3)	7 (29.2)	8 (38.1)	7 (29.2)	15 (32.6)
Never	3 (14.3)	5 (20.8)	2 (9.5)	5 (20.8)	8 (17.4)
Current	11 (52.4)	12 (50.0)	11 (52.4)	12 (50.0)	23 (50.0)
Receiving ART					
Yes	20 (95.2)	22 (91.7)	20 (95.2)	22 (91.7)	43 (93.5)
No	1 (4.8)	2 (8.3)	1 (4.8)	2 (8.3)	3 (6.5)
Pack-years					
Mean±SD	22.0±17.3	14.3±13.6	25.8±20.7	14.3±13.6	19.2±18.1
Median (min, max)	20.0 (0, 56.0)	12.1 (0, 38.0)	20.5 (0, 80.0)	12.1 (0, 38.0)	15.0 (0, 80.0)
FEV_1 % pred[#]					
Mean±SD	68.3±16.8	104±11.2	67.7±17.1	104±11.2	86.9±23.2
Median (min, max)	68.3 (21.0, 90.4)	105 (81.3, 128)	68.3 (21.0, 90.4)	105 (81.3, 128)	88.2 (21.0, 128)
FEV_1[#]					
Mean±SD	2.07±0.603	3.32±0.737	2.04±0.608	3.32±0.737	2.71±0.925
Median (min, max)	2.18 (0.650, 3.29)	3.21 (1.95, 4.77)	2.00 (0.650, 3.29)	3.21 (1.95, 4.77)	2.61 (0.650, 4.77)
FEV_1/FVC[#]					
Mean±SD	0.560±0.118	0.794±0.0590	0.556±0.119	0.794±0.0590	0.681±0.149
Median (min, max)	0.592 (0.293, 0.679)	0.783 (0.689, 0.905)	0.588 (0.293, 0.679)	0.783 (0.689, 0.905)	0.692 (0.293, 0.905)
D_{LCO} % pred					
Mean±SD	73.1±26.7	74.6±22.3	71.9±27.1	74.6±22.3	73.9±24.1
Median (min, max)	78.0 (36.3, 139)	72.1 (14.4, 116)	67.3 (36.3, 139)	72.1 (14.4, 116)	73.3 (14.4, 139)
Missing	4 (19.0)	5 (20.8)	5 (23.8)	5 (20.8)	10 (21.7)
Viral load					
<50 copies	11 (52.4)	16 (66.7)	10 (47.6)	16 (66.7)	27 (58.7)
>50 copies	1 (4.8)	3 (12.5)	1 (4.8)	3 (12.5)	4 (8.7)
Missing	9 (42.9)	5 (20.8)	10 (47.6)	5 (20.8)	15 (32.6)

Data are presented as n (%) unless indicated otherwise. Those with and without OLD were matched based on age, smoking status and antiretroviral treatment (ART) status. TMT: tandem mass tagging; BALF: bronchoalveolar lavage fluid; OLD: obstructive lung disease; FEV_1: forced expiratory volume in 1 s; FVC: forced vital capacity; D_{LCO}: diffusing capacity of the lung for carbon monoxide. [#]: significant difference between those with and without OLD (using all available samples) at the 0.05 level.

(71.7%) with a mean age of 56.3 years, and 54.3% identified as black, non-Hispanic, 43.5% as white or Hispanic/Latino, and 2.2% identified as Asian or Pacific Islander. Most participants were receiving antiretroviral treatment (93.5%) at the time of study. Smoking status was similar between those with and without OLD, with 50% actively smoking at the time of enrolment; however, pack-years were greater in those with OLD (22.0–25.8) *versus* those without OLD (14.3). Lung function (FEV_1) ranged from 21 to 128% of predicted normal. Among those with OLD, the mean FEV_1 % pred was 67.7%, whereas in those without OLD the mean was 104%.

Proteome associations with lung function

We measured 1219 unique proteins in the soluble component and 2628 proteins in the insoluble component of BALF using a non-targeted TMT labelling and LC-MS analysis approach, with 645 proteins common to both sample types (figure 1a). We performed both unadjusted and adjusted for smoking status analysis. In our unadjusted analysis we identified 81 proteins in the soluble component and 61 proteins in the insoluble BALF component that correlated with FEV_1 % pred (FDR <0.1). The top 10 proteins that most significantly associated with FEV_1 % pred are depicted in supplementary table 1S. However, when we adjusted for smoking, we identified only a single protein, mucin-5AC, in the soluble component and no proteins in the insoluble components correlated with FEV_1 % pred at an FDR $\leq$0.05 (supplementary table 2S). Validation of mucin-5AC was performed using PRM analysis on the original cohort. Four peptides specific for mucin-5AC were identified and quantified in 10 samples, five from individuals with low FEV_1 % pred and five from those with high FEV_1 % pred. PRM confirmed the presence of mucin-5AC and quantitation was consistent with the MS data (supplementary table 3S).

Combined proteomic analysis

To enhance our proteomic coverage we included in our analysis the previously reported SomaScan targeted proteomic data from the BALF soluble component using a permutation testing framework [14]. The SomaScan proteome provided an additional 3872 proteins, of which 1122 overlapped with proteins measured by TMT. We first describe results without an adjustment for smoking status. There were 401 proteins showing significant associations with FEV_1 % pred (including the soluble component and insoluble component TMT proteins and SomaScan) after permutation testing and multiple comparisons adjustment at an FDR level of 0.05. Out of these 401 proteins, 98 were detected in the soluble component using TMT, 123 in the insoluble component using TMT, and 342 in the soluble component using

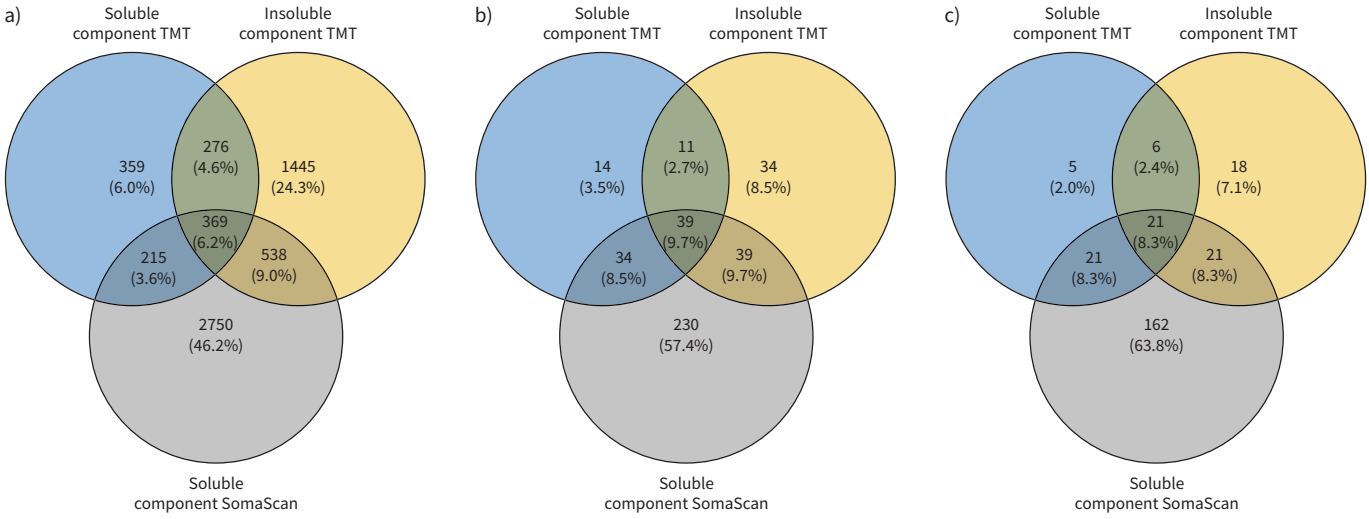

FIGURE 1 a) Venn diagram of the overlap in proteins measured in each BALF component using either TMT (soluble component or insoluble component) or SomaScan (soluble component). b) Venn diagram of the overlap in proteins measured in each BALF component using either TMT (soluble component or insoluble component) or SomaScan (soluble component) that were also significantly correlated with FEV_1 % pred prior to smoking adjustment and after multiple comparisons at an FDR 0.05 level. Correlation with FEV_1 % pred was assessed using a permutation testing framework to use information across all three datasets. c) Venn diagram of the overlap in proteins measured in each BALF component using either TMT (soluble component or insoluble component) or SomaScan (soluble component) that were also significantly correlated with FEV_1 % pred after smoking adjustment and after multiple comparisons at an FDR 0.05 level. Correlation with FEV_1 % pred was assessed using a permutation testing framework to use information across all three datasets. TMT: tandem mass tagging; BALF: bronchoalveolar lavage fluid; FEV_1: forced expiratory volume in 1 s; FDR: false discovery rate.

TABLE 2 Top pathways based on the proteins significant prior to smoking adjustment at an FDR level of 0.05 after permutation testing using BALF soluble component TMT, BALF insoluble component TMT and soluble component SomaScan datasets

Pathway	Source	No. of overlapping genes	No. of all pathway genes	p-value	q-value
Cell surface interactions at the vascular wall	Reactome	22	100 (215)	2.20e-06	0.0102
Cell adhesion molecules: *Homo sapiens* (human)	KEGG	17	76 (149)	2.54e-05	0.0595

FDR: false discovery rate; TMT: tandem mass tagging; BALF: bronchoalveolar lavage fluid.

SomaScan. There were 39 common proteins with significant associations with FEV_1 % pred in the soluble component using TMT and SomaScan as well as the insoluble component using TMT (figure 1b). None of the 401 proteins were significantly correlated with D_{LCO} % pred. We used these 401 proteins in a pathway analysis which revealed a single pathway that was significant after multiple comparisons adjustment at the FDR 0.05 level (p=0.01) and a second pathway significant after adjustment at the FDR 0.1 level (p=0.06) The most-significant pathway was cell surface interactions at the vascular wall followed by cell adhesion molecules (table 2).

We applied the same permutation testing framework with an adjustment for smoking status. We found 254 proteins significant at the FDR level of 0.05. Of these, there were 53 proteins from the soluble component detected by TMT, 66 from the insoluble component detected by TMT and 225 from the soluble component detected by SomaScan (figure 1c). None of these proteins were significantly associated with D_{LCO} % pred. There were 240 proteins that were significantly correlated with FEV_1 % pred both with and without adjusting for smoking status. There were 21 proteins with significant associations with FEV_1 % pred detected in all three datasets. Using these 254 proteins we identified three significant pathways at the FDR level of 0.05. They included axon guidance, nervous system development and cell adhesion molecules (table 3).

To identify the types of proteins residing in each BALF compartment, we compared the three ontology processes cellular component, biological process and molecular function of the soluble component TMT, insoluble component TMT and soluble component SomaScan proteins (supplementary figure 1S). The soluble component TMT and SomaScan showed significant overlap, while the insoluble component TMT proteins differed significantly. The proteins identified within the insoluble component represented the membrane, vesicle and vacuole cellular components with biological and molecular functions representing protein targeting, GTP metabolism and protein translation, respectively. For proteins that associated with FEV_1 % pred, the insoluble component proteins were associated with vesicles, and functions related to cellular adhesion.

Endotypes identified by cluster analysis

We visualised the protein abundances in the BALF soluble component TMT and insoluble component TMT using heatmaps, which revealed a distinct subgroup of patients with lower lung function (FEV_1 % pred) within the soluble component proteome (figure 2a), but not with the insoluble component proteome (figure 2b). We constructed protein-driven endotypes using K-means clustering for both the proteome in

TABLE 3 Top pathways based on the proteins significant after smoking adjustment at an FDR level of 0.05 after permutation testing using BALF soluble component TMT, BALF insoluble component TMT and soluble component SomaScan datasets

Pathway	Source	No. of overlapping genes	No. of all pathway genes	p-value	q-value
Axon guidance	Reactome	22	182 (356)	1.82e-05	0.0439
Nervous system development	Reactome	22	187 (379)	2.80e-05	0.0439
Cell adhesion molecules: *Homo sapiens* (human)	KEGG	13	76 (149)	2.82e-05	0.0439

FDR: false discovery rate; TMT: tandem mass tagging; BALF: bronchoalveolar lavage fluid.

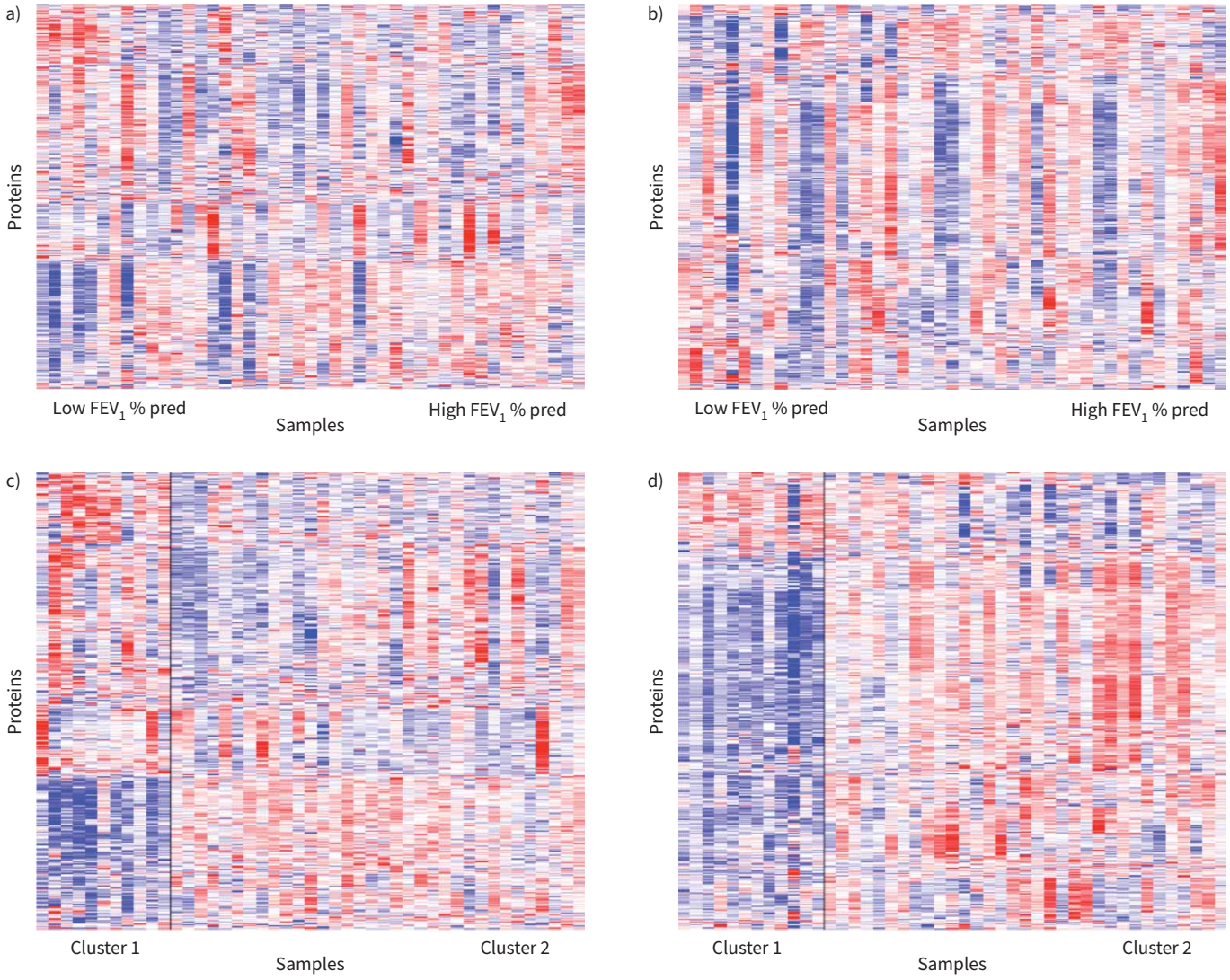

FIGURE 2 Heatmaps of protein abundances in BALF soluble component and insoluble component TMT datasets. Proteins are ordered using hierarchical clustering with average linkage. a) Protein abundances in the BALF soluble component TMT dataset with samples ordered from lowest to highest FEV_1 % pred. b) Protein abundances in the BALF insoluble component TMT dataset with samples ordered from lowest to highest FEV_1 % pred. c) Protein abundances in the BALF soluble component TMT dataset with samples grouped using K-means clustering. d) Protein abundances in the BALF insoluble component TMT dataset with samples grouped by K-means clustering. TMT: tandem mass tagging; BALF: bronchoalveolar lavage fluid; FEV_1: forced expiratory volume in 1 s.

the soluble component and insoluble component (figure 2c,d) using all identified proteins. Fixing K=2, we identified two clusters on the soluble component TMT and insoluble component TMT datasets. The soluble component Cluster 1 was similar to that previously reported in the BALF soluble component SomaScan proteome and metabolome [14]. Cluster 1 largely comprised individuals who identified as black (81.8%) compared to Cluster 2 where 44.1% of individuals identified as black. Cluster 1 also comprised individuals with lower mean FEV_1 % pred (67.7 *versus* 94.0), lower mean FEV_1/FVC (0.523 *versus* 0.737) and 81.8% of individuals had OLD compared to 35.3% in Cluster 2 (supplementary table 4S). The insoluble component clusters differed significantly from the soluble component clusters. For the insoluble component the two clusters, Clusters 1 and 2 respectively, were similar in the percentage of patients with OLD (41.7% *versus* 48.5%), mean age (56.8 *versus* 55.9), current smokers (50.0% *versus* 51.5%), mean FEV_1 % pred (86.8% *versus* 87.5%) and mean FEV_1/FVC (0.751 *versus* 0.658). The insoluble component proteome clusters were most distinct in terms of ethnicity: 33.3% of patients in Cluster 1 identified as black, non-Hispanic compared to 60.6% in Cluster 2 and 66.7% identified as white, non-Hispanic compared to 36.4% in Cluster 2 (supplementary table 5S).

We then considered BCC to identify an overall clustering scheme on the BALF soluble component TMT, insoluble component TMT and SomaScan proteome datasets based on individual-level clustering. The individual-level soluble component TMT and SomaScan clusters were demographically similar: Cluster 1 (the smaller cluster) based on either dataset is similar in the percentage of male participants (BALF soluble component TMT: 100%, SomaScan: 77.8%), mean age (BALF soluble component TMT: 63.0, SomaScan: 61.8), and mean FEV_1 % pred (BALF soluble component TMT: 67.7%, SomaScan: 64.8%). The insoluble component TMT clusters, however, have demographic features that differ. For the insoluble component, Cluster 1 (the smaller cluster) is 60% male with an average age of 54.3years and with mean FEV_1 % pred as 89.7%. The overall clustering scheme is more like the soluble component TMT and SomaScan clusters with Cluster 1 being 100% male, slightly older (mean age=63.0), with lower FEV_1 % pred (mean=67.7%) compared to Cluster 2 (supplementary table 6S). The similarity between the clusters determined using the BALF soluble component TMT and SomaScan datasets with the overall clustering scheme aligns with the high degree of correlation between proteins measured using TMT in the soluble component and the same proteins measured using SomaScan in the soluble component (figure 3a). In contrast, the correlations between proteins measured in the BALF insoluble component using TMT and proteins measured in the soluble component using SomaScan were generally lower (figure 3b).

Discussion

We performed a broad proteomic analysis on BALF from PLWH to identify proteins and biological pathways that associate with OLD. For this study we primarily considered FEV_1 % pred as an outcome as our previous work indicated that there was considerable heterogeneity in lung function, as measured by FEV_1 % pred, within those with and without OLD. Approximately a third of our participants smoked tobacco at the time of BALF sampling. To compare the effects of smoking with that of HIV, we included results both adjusted and unadjusted for smoking status.

BALF is a complex mixture of proteins, metabolites and lipids. BALF preparation techniques typically involve centrifugation to remove cells, often followed by additional filtering or centrifugation to separate out other debris that is discarded, which tends to be lipid-rich components. All the samples had cells removed prior to freezing and storage. We then performed a two-step centrifugation process to collect debris separately from the soluble component and performed untargeted MS-based proteomics on both the soluble component and the cell-free insoluble component to expand our proteome coverage. This broadened our proteomic findings as the insoluble component and soluble component represented different cellular components that differed in both biological and molecular processes. We found that both the soluble component and the insoluble component proteomes were associated with FEV_1 % pred among PLWH.

Using an untargeted, TMT labelling and LC-MS approach to identify and quantify proteins, we found a number of proteins that correlated with FEV_1 % pred in our unadjusted model, but only a single protein,

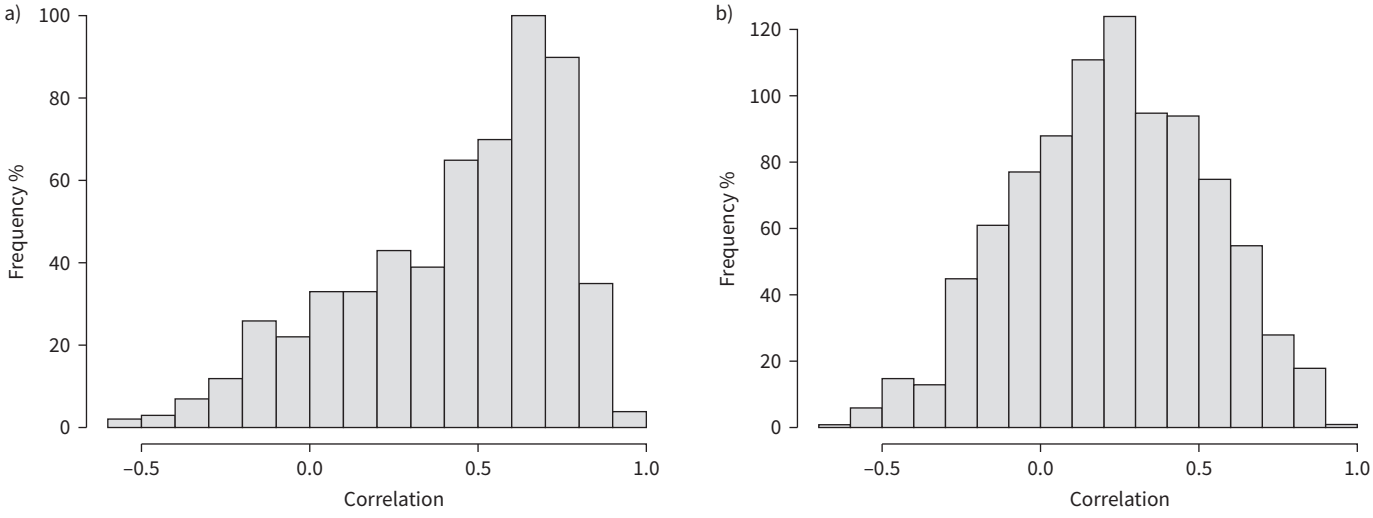

FIGURE 3 a) Histogram of the distribution of correlations between proteins measured in the BALF soluble component using TMT and SomaScan. b) Histogram of the distribution of correlations between proteins measured in the BALF insoluble component using TMT and in the soluble component using SomaScan. TMT: tandem mass tagging; BALF: bronchoalveolar lavage fluid.

mucin-5AC, reached the FDR threshold in the adjusted models. This suggests that proteins represented in the TMT proteome were largely influenced by smoking status. Most of these proteins across the two TMT datasets did not overlap, consistent with these separate BALF fractions representing different components with differing biological and molecular functions. The notion that the soluble component and insoluble component proteins capture different aspects of activity in BALF was emphasised by differing levels of correlation between overlapping proteins measured by both SomaScan and TMT in the BALF and insoluble component (supplementary figure S3).

One significant protein from the soluble component, irrespective of smoking status, included mucin-5AC, a glycoprotein that protects the airway against infection. Elevated levels of mucin-5AC have been associated with airway hypersecretion, particularly in COPD patients with predominantly airway disease [22, 23]. SP-A2 was increased in both the adjusted and unadjusted models; however, it only reached significance in the unadjusted model. SP-A2, a collectin, is an abundant protein that makes up pulmonary surfactant, an insoluble complex of proteins and phospholipids. Surfactant A is localised to several cells in the lung, including type II pneumocytes, goblet cells and club cells, and elevated levels have been associated with COPD [24]. We were able to link significant proteins to biological pathways, but only in the insoluble component fraction. These included cell adhesion and haematopoietic cell lineage pathways, with the former implicated in COPD and lung inflammation [25–27].

We examined associations between the lung proteome and lung function (FEV_1 % pred) by uniquely combining our untargeted proteomics data from BALF soluble component and BALF insoluble component with a targeted proteomics assay from BALF soluble component using SomaScan technology. This broadened our depth of coverage for the proteome as most proteins in each group were unique. Our permutation-based approach to combine results across three datasets yielded 401 proteins significantly correlated with FEV_1 % pred when unadjusted for smoking, and after multiple comparisons adjustment we identified a significant pathway. This pathway, cell surface interactions at the vascular wall, has been previously reported as an enriched pathway among COPD-omics subtypes [28]. Vascular remodelling is associated with lung decline in COPD and may contribute to gas exchange impairment and loss of the alveolar structure [29]. This pathway may play a role in the reduction in pulmonary diffusing capacity and increased presence of emphysema in PLWH infection who are actively smoking [30–32]. In our analysis adjusted for smoking, we identified three pathways, one of which, cell adhesion molecules, was significant in both unadjusted and adjusted models. Cell adhesion molecular families are involved in the inflammatory response in the lung and are responsible for both epithelial and endothelial integrity [33]. These molecules have recently been identified as therapeutic targets for COPD, particularly COPD exacerbations, and appear to be prominent irrespective of smoking status [26].

Previously, we reported on an endotype driven by unique proteomic and metabolomic abundance patterns [14]. This endotype was generally composed of older men with especially poor lung function who largely identified as black, non-Hispanic. The proteins associated with this cluster mapped to pathways involved with insulin, FOXO transcription factors, apoptosis, RNA metabolism and retinol metabolism. We corroborated this endotype in the BALF soluble component proteome using an untargeted TMT labelled MS approach. Consensus clustering revealed that analysis of the BALF soluble component by TMT and SomaScan adhered more closely to a similar clustering pattern that was not apparent from the insoluble component fraction. The correlation between the soluble component TMT and SomaScan likely relates to the testing of the same BALF compartment, *i.e.* the BALF soluble component and its associated proteome. We found that although the insoluble component TMT proteome differentiated patients with OLD, its associated cluster did not have an identifiable demographic or physiological, *i.e.* lung function (FEV_1 % pred), association. This difference in clustering likely relates to the insoluble component and soluble component differing in the cellular compartment and biological and molecular processes that they represented.

There are several limitations to this study. The small sample size of our data likely limited the number of pathways we could detect among the significant proteins. A larger sample size may give us more power to identify proteins and pathways associated with OLD or COPD pathogenesis. In addition, there are many confounders that could impact our findings, such as previous lung infection, environmental exposures and other medications that we were unable to account for. Lastly, our study also did not include HIV-negative controls, limiting our understanding of the involvement of the HIV infection itself in our results.

Overall, the strength of this study is the depth of proteome coverage and the combination of complimentary proteomic platforms such as an MS-based untargeted approach and the targeted approach of SomaScan. In addition, by comparing unadjusted to adjusted models for smoking, we can compare the effects of smoking status on the lung proteome.

In conclusion, we were able to utilise a combination of untargeted and targeted proteomic approaches to identify proteins and pathways that associated with HIV-associated OLD. This combination greatly expanded the proteomic coverage of the BALF. In addition, we validated a unique endotype consisting primarily of black men with severe obstruction with a unique proteomic and metabolomic signature. Furthermore, we demonstrated that the lipid-rich insoluble component, which often is discarded prior to processing, identifies proteins unique from the BALF soluble component and ones that correlate with lung function.

Provenance: Submitted article, peer reviewed.

Acknowledgements: We thank the Center for Metabolomics and Proteomics at the University of Minnesota for providing services related to the generation of quantitative proteomics and metabolomics.

Author contributions: Conception and design: S. Samorodnitsky, E.F. Lock, K.M. Kunisaki, L. Parker, P. Jagtap, T.J. Griffin and C.H. Wendt. Data acquisition: D. Weise, M. Kruk, L. Parker, P. Jagtap, T.J. Griffin and C.H. Wendt. Analysis and interpretation: S. Samorodnitsky, E.F. Lock, K.M. Kunisaki, A. Morris, J.M. Leing, P. Jagtap, T.J. Griffin and C.H. Wendt. Drafting of manuscript: S. Samorodnitsky, D. Weise and C.H. Wendt. Editing of manuscript: S. Samorodnitsky, D. Weise, E.F. Lock, K.M. Kunisaki, A. Morris, J.M. Leung, L. Parker, P. Jagtap, T.J. Griffin and C.H. Wendt.

Conflict of interest: S. Samorodnitsky reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1. D. Weise reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1; and support for attending meetings and/or travel from National Institutes of Health grant R01 HL140971-01A, outside the submitted work. E.F. Lock reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1. K.M. Kunisaki reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1; consulting fees from Allergan, outside the submitted work; participation in data safety monitoring board for Nuvaira and Organicell, outside the submitted work; and is a volunteer for the American Thoracic Society Proposal Review Committee, outside the submitted work. A. Morris reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1. J.M. Leung reports support for the present manuscript from the Canadian Institutes of Health Research and BC Lung Foundation; grants or contracts from the Canadian Institutes of Health Research and BC Lung Foundation, outside the submitted work; payment received from the BC Lung Foundation for patient forums, outside the submitted work; payment received from the University of British Columbia for education lectures, outside the submitted work; participation in the Enhance Quality Safety and Patient Experience in Chronic Obstructive Pulmonary Disorder Data and Safety Monitoring Board, outside the submitted work; and is a member of the steering committee for Canadian Respiratory Research Network and CanCOLD Study, outside the submitted work. M. Kruk reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1; and support for attending meetings and/or travel from National Institutes of Health grant R01 HL140971-01A, outside the submitted work. L. Parker reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1. P. Jagtap reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1. T.J. Griffin reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1. C.H. Wendt reports support for the present manuscript from National Institutes of Health grant R01 HL140971-01A1; support for attending meetings and/or travel from National Institutes of Health grant R01 HL140971-01A, outside the submitted work; and participation on a data and safety monitoring board for the Alpha1 Biomarkers Consortium and NIH, outside the submitted work.

Support statement: Supported by National Institutes of Health grant R01 HL140971-01A1 (all authors). The Orbitrap Eclipse instrumentation platform used in this work was purchased through High-end Instrumentation Grant S10OD028717 from the NIH. This material is also the result of work supported with resources and the use of facilities at the Minneapolis Veterans Affairs Medical Center, Minneapolis, MN, USA. The views expressed in this article are those of the authors and do not reflect the views of the US Government, the Department of Veterans Affairs, the funders, the sponsors or any of the authors' affiliated academic institutions. Funding information for this article has been deposited with the Crossref Funder Registry.

References

1 Drummond MB, Merlo CA, Astemborski J, *et al.* The effect of HIV infection on longitudinal lung function decline among IDUs: a prospective cohort. *AIDS* 2013; 27: 1303–1311.
2 Gingo MR, George MP, Kessinger CJ, *et al.* Pulmonary function abnormalities in HIV-infected patients during the current antiretroviral therapy era. *Am J Respir Crit Care Med* 2010; 182: 790–796.
3 Hirani A, Cavallazzi R, Vasu T, *et al.* Prevalence of obstructive lung disease in HIV population: a cross sectional study. *Respir Med* 2011; 105: 1655–1661.

4 Kristoffersen US, Lebech AM, Mortensen J, *et al.* Changes in lung function of HIV-infected patients: a 4.5-year follow-up study. *Clin Physiol Funct Imaging* 2012; 32: 288–295.
5 Crothers K, Huang L, Goulet JL, *et al.* HIV infection and risk for incident pulmonary diseases in the combination antiretroviral therapy era. *Am J Respir Crit Care Med* 2011; 183: 388–395.
6 Cui Q, Carruthers S, McIvor A, *et al.* Effect of smoking on lung function, respiratory symptoms and respiratory diseases amongst HIV-positive subjects: a cross-sectional study. *AIDS Res Ther* 2010; 7: 6.
7 Madeddu G, Fois AG, Calia GM, *et al.* Chronic obstructive pulmonary disease: an emerging comorbidity in HIV-infected patients in the HAART era? *Infection* 2013; 41: 347–353.
8 Kunisaki KM, Niewoehner DE, Collins G, *et al.* Pulmonary function in an international sample of HIV-positive, treatment-naive adults with CD4 counts >500 cells/muL: a substudy of the INSIGHT Strategic Timing of AntiRetroviral Treatment (START) trial. *HIV Med* 2015; 16: Suppl. 1, 119–128.
9 Morris A, Sciurba FC, Lebedeva IP, *et al.* Association of chronic obstructive pulmonary disease severity and *Pneumocystis* colonization. *Am J Respir Crit Care Med* 2004; 170: 408–413.
10 Fitzpatrick ME, Singh V, Bertolet M, *et al.* Relationships of pulmonary function, inflammation, and T-cell activation and senescence in an HIV-infected cohort. *AIDS* 2014; 28: 2505–2515.
11 Qin S, Vodovotz L, Zamora R, *et al.* Association between inflammatory pathways and phenotypes of pulmonary dysfunction using cluster analysis in persons living with HIV and HIV-uninfected individuals. *J Acquir Immune Defic Syndr* 2020; 83: 189–196.
12 Jan AK, Moore JV, Wang RJ, *et al.* Markers of inflammation and immune activation are associated with lung function in a multi-center cohort of persons with HIV. *AIDS* 2021; 35: 1031–1040.
13 Lawani MB, Morris A. The respiratory microbiome of HIV-infected individuals. *Expert Rev Anti Infect Ther* 2016; 14: 719–729.
14 Samorodnitsky S, Lock EF, Kruk M, *et al.* Lung proteome and metabolome endotype in HIV-associated obstructive lung disease. *ERJ Open Res* 2023; 9: 00332-2022.
15 Cribbs SK, Uppal K, Li S, *et al.* Correlation of the lung microbiota with metabolic profiles in bronchoalveolar lavage fluid in HIV infection. *Microbiome* 2016; 4: 3.
16 Akata K, Leung JM, Yamasaki K, *et al.* Altered polarization and impaired phagocytic activity of lung macrophages in people with HIV and COPD. *J Infect Dis* 2022; 225: 862–867.
17 Weise DO, Kruk ME, Higgins L, *et al.* An optimized workflow for MS-based quantitative proteomics of challenging clinical bronchoalveolar lavage fluid (BALF) samples. *Clin Proteomics* 2023; 20: 14.
18 Benjamini YYH. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc* 1995; 57: 289–300.
19 Kamburov A, Cavill R, Ebbels TM, *et al.* Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics* 2011; 27: 2917–2918.
20 Elson RC. On Fisher's method of combining p-values. *Biom J* 1991; 33: 339–345.
21 Lock EF, Dunson DB. Bayesian consensus clustering. *Bioinformatics* 2013; 29: 2610–2616.
22 Huang JT, Cant E, Keir HR, *et al.* Endotyping chronic obstructive pulmonary disease, bronchiectasis, and the "Chronic obstructive pulmonary disease-bronchiectasis association". *Am J Respir Crit Care Med* 2022; 206: 417–426.
23 Dasgupta A, Chakraborty R, Saha B, *et al.* Sputum protein biomarkers in airway diseases: a pilot study. *Int J Chron Obstruct Pulmon Dis* 2021; 16: 2203–2215.
24 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: 5125–5132.
25 Riise GC, Larsson S, Lofdahl CG, *et al.* Circulating cell adhesion molecules in bronchial lavage and serum in COPD patients with chronic bronchitis. *Eur Respir J* 1994; 7: 1673–1677.
26 Vanderslice P, Biediger RJ, Woodside DG, *et al.* Development of cell adhesion molecule antagonists as therapeutics for asthma and COPD. *Pulm Pharmacol Ther* 2004; 17: 1–10.
27 Woodside DG, Vanderslice P. Cell adhesion antagonists: therapeutic potential in asthma and chronic obstructive pulmonary disease. *BioDrugs* 2008; 22: 85–100.
28 Hobbs BD, Morrow JD, Wang XW, *et al.* Identifying chronic obstructive pulmonary disease from integrative omics and clustering in lung tissue. *BMC Pulm Med* 2023; 23: 115.
29 Harkness LM, Kanabar V, Sharma HS, *et al.* Pulmonary vascular changes in asthma and COPD. *Pulm Pharmacol Ther* 2014; 29: 144–155.
30 Attia EF, Akgun KM, Wongtrakool C, *et al.* Increased risk of radiographic emphysema in HIV is associated with elevated soluble CD14 and nadir CD4. *Chest* 2014; 146: 1543–1553.
31 Ronit A, Kristensen T, Hoseth VS, *et al.* Copenhagen Comorbidity in HIVIsg. Computed tomography quantification of emphysema in people living with HIV and uninfected controls. *Eur Respir J* 2018; 52: 1800296.
32 Crothers K, McGinnis K, Kleerup E, *et al.* HIV infection is associated with reduced pulmonary diffusing capacity. *J Acquir Immune Defic Syndr* 2013; 64: 271–278.
33 Pilewski JM, Albelda SM. Adhesion molecules in the lung. An overview. *Am Rev Respir Dis* 1993; 148: S31–S37.