

BAL Fluid Metaproteome in Acute Respiratory Failure

To the Editor:

Studies have shown that lung microbiota are altered from the healthy state during respiratory disease. Gaining insights into the interactions between the host and the microbiome is of paramount importance to increase our understanding of both health and disease. A limitation of gene marker analysis to infer phylogenetic information about microbial communities (1, 2) is that sequenced DNA could originate from inactive cells. Recent techniques such as presorting live microbes (3), stable-isotope probing (4), and RNA sequencing (5) have emerged as tools for providing insights into active community members. Metaproteomics detects proteins expressed by bacterial communities as potential markers of microbial activity (6). Given the limitations of DNA sequencing and culture-dependent methods for identifying respiratory pathogens, we sought to develop a mass spectrometry (MS)-based proteomics and bioinformatics pipeline for determining the composition of active lung microbial communities.

We previously reported findings from quantitative studies of BAL fluid (BALF) from patients with lung injury after hematopoietic stem cell transplantation (idiopathic pneumonia syndrome, $n = 12$; infectious lung injury, $n = 18$) (7) or acute respiratory distress syndrome ($n = 36$) (8). The majority of the patients had widespread lung damage and were on antibiotics. These studies used isobaric tagging to quantify host human proteins. For the proof-of-concept study described here, we reanalyzed this dataset using spectral counting to provide an estimate of microbial peptide abundance. The spectral datasets for hematopoietic stem cell transplantation (8) and acute respiratory distress syndrome (9) were combined for our analysis. The RAW files containing combined spectrum-level data were searched against a custom protein sequence database (described below). We processed 173 MS RAW files containing 926,102 tandem mass spectra (MS/MS) using sequential analytical workflows within the Galaxy-P platform (Figure 1). The raw data for the analyses of this dataset have been deposited in the ProteomeXchange Consortium with the dataset identifier PXD008273.

Workflow 1

We created a protein sequence database containing 2,837,537 human and microbial peptides expected in BALF by merging five publically available databases: 1) the Human UniProt database, 2) contaminant proteins from the common Repository of Adventitious Proteins, 3) the Oral Human Microbiome Project Reference Genome Sequence

Data database, 4) the Human Oral Microbiome database, and 5) the Human Microbiome Project gastrointestinal tract database. This workflow created peaklists from the MS/MS data for matching to the customized protein sequence database. Controlling for a false discovery rate of $\leq 5\%$, 235,602 peptide MS/MS spectra matched to peptide sequences in this database.

Workflow 2

Peptide-spectrum matches (PSMs) from workflow 1 that did not match Human UniProt, decoy, or contaminant species were used to identify microbial sequences. These PSMs were analyzed with Megan6 to extract phylogenetic and functional information (9). We identified 776 microbial PSMs (426 distinct bacterial peptide sequences) that were further examined with Megan6 (Table E1 in the data supplement).

We identified 63 genera (Figure E1), 41 of which were reported in at least one prior metagenome study in sputum or BALF (Table 1). Twenty of the 27 genera that were identified with at least two peptides were also reported in prior metagenomics studies (Table 1). These included several known pulmonary pathogenic genera, including *Burkholderia*, *Klebsiella*, *Listeria*, and *Staphylococcus*. Clinical records for these patients identified normal respiratory flora and six other microorganisms, four of which were detected by our metaproteomic approach. Nine genera, including *Pyramidobacter*, *Actinomyces*, and *Streptococcus*, were detected from at least 20 clinical samples (Table 1). These microbial peptides mapped to several Gene Ontology annotations, including regulation of transcription, protein translation, trafficking, protein folding, nucleic acid metabolism, amino acid metabolism, carbohydrate metabolism, glycolysis, gluconeogenesis, fatty acid oxidation, redox response, cellular response to stress, and cell adhesion (Figure E2).

There is growing evidence of a complex interplay between respiratory microbes and host in both health and disease. Thus far, most of the studies evaluating microbial diversity in the lung have used marker gene analysis of 16S rRNA-encoding genes (1, 10–12), which does not reveal the host–microbe interactions. Our pipeline demonstrates the feasibility of identifying microbial peptides, bacterial taxa, and associated biological functions activated by the microbial community in the distal lung.

Importantly, a functional analysis may suggest that the microbiota we identified were metabolically active (Table E1, Figure E3). Many of the proteins we detected are known to be involved in energy production, protein transcription, or protein assembly. For instance, the most abundant protein recovered, triose-phosphate isomerase, is a key enzyme in the glycolytic pathway. We also discovered members of the heat shock protein 70 family, whose constituents chaperone protein folding. Protein regulators of transcription, including transcription factor Helix-turn-helix (HTH) DNA binding motif and members of the Repressor, Orf, Kinase (ROK) family, were also recovered.

Although our pipeline serves as an important proof of concept, we acknowledge several key areas in need of improvement. To increase detection of microbial proteins within a host sample, enrichment of BALF microbiota and/or sample fractionation coupled with the use of increasingly sensitive mass spectrometers could be pursued. To improve our coverage of lower-abundance proteins, we depleted medium- and high-abundance proteins (13)—a strategy that could remove other proteins bound to the high-abundance

Supported by an American Heart Association Scientist Development Grant (12SDG8160000) (M.B.) and the University of Minnesota CTSI KL2 Scholars Program (KL2 RR0333182) (M.B.). The Galaxy-P project is supported by National Science Foundation grant DBI-1458524 (T.J.G.) and National Institutes of Health grant U24CA199347 (T.J.G.).

Author Contributions: Study conception and design: P.D.J., K.J.V., and M.B. Acquisition of data: P.D.J., K.J.V., and M.B. Analysis and interpretation of data: P.D.J., K.J.V., J.J., T.M., K.M.P., T.J.G., R.C.H., J.D.R., and M.B. Drafting or revision of the manuscript: P.D.J., K.J.V., J.J., T.M., K.M.P., T.J.G., R.C.H., J.D.R., and M.B. Final approval of the manuscript: P.D.J., K.J.V., J.J., T.M., K.M.P., T.J.G., R.C.H., J.D.R., and M.B.

This letter has a data supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

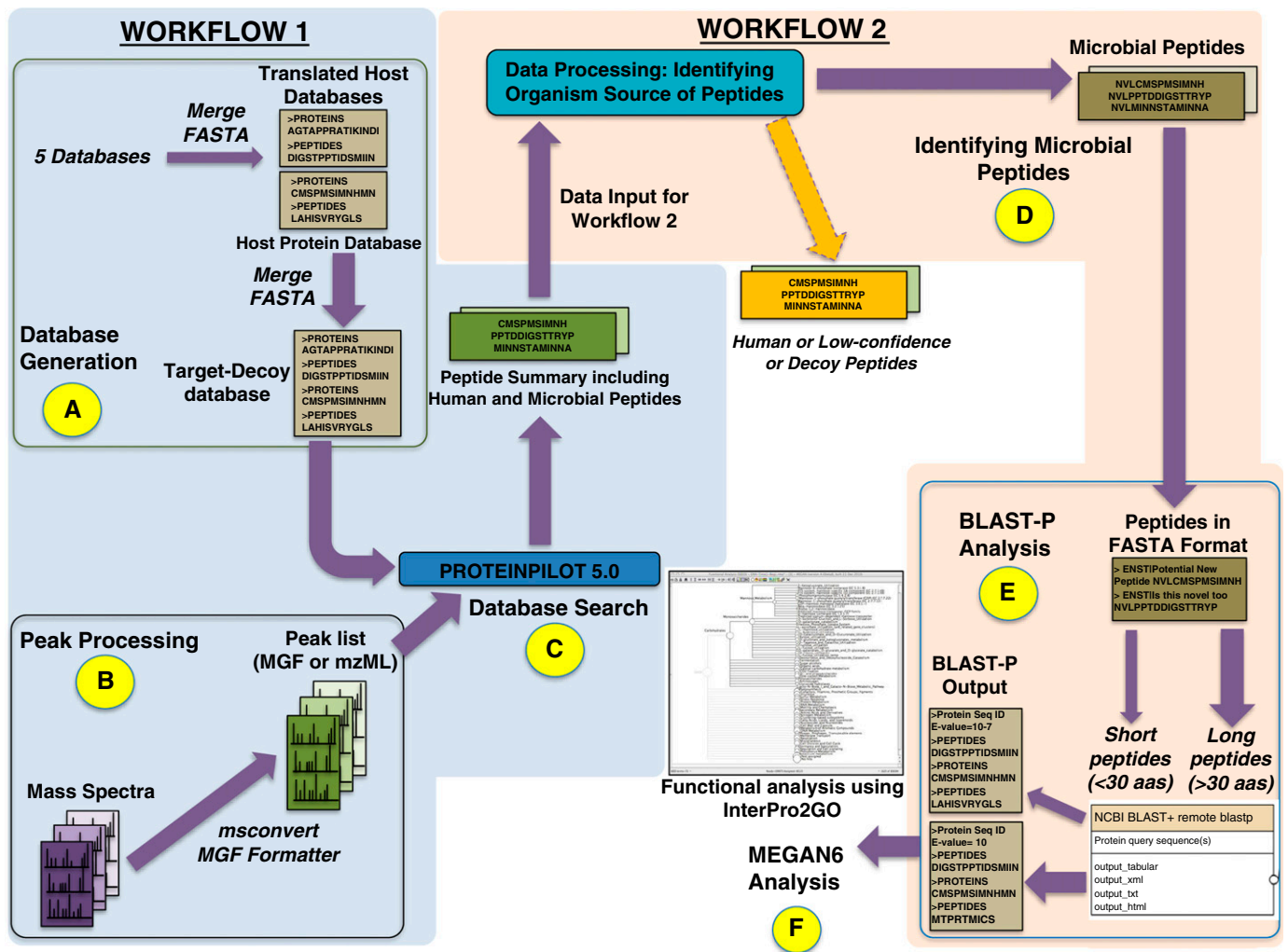


Figure 1. Metaproteomics workflow in Galaxy-P to identify bacterial peptides in BAL fluid. In the first workflow (steps A–C), a database was generated (A) and peak processing RAW files were converted to mzml and MGF files and searched with ProteinPilot version 5.0.0, 1654 revision: 1656. (B). The outputs from these steps were used for a database search (C) to generate a list of both human and bacterial peptide-spectrum matches (PSMs) (Table E1). In the second workflow (steps D–F), bacterial PSMs were parsed out by eliminating human-origin peptides from the list (D). The microbial peptides were subjected to BLAST-P analysis (E) against the NCBI nr database to generate an output for subsequent MEGAN6 analysis. MEGAN6 software uses a bitscore threshold to assign matches. Any match that has a bitscore below the threshold is not used to assign taxonomy. If all matches for a taxonomy or function are below the threshold, it is marked as unassigned. Peptides that do not have a match because of an absence of sequences in the InterPro2GO mapping file are termed “No Hits” (F).

proteins. Another limitation of the field is the absence of databases of respiratory microbial peptides for spectral matching. A high-quality metagenomic sequencing database is critical, as peptide sequence identification relies on the fidelity of predicted protein sequences. An approach using simultaneous whole-metagenome sequencing from the same sample could substantially increase the detection of bacterial peptides for metaproteomic studies. Such advances could propel metaproteomics to a point where it rivals the depth of microbial characterization offered by 16S rRNA sequencing, with the added benefit of providing a functional snapshot of the system. Although the use of a pooled spectral dataset does not allow absolute quantification of bacterial peptides from individual patients, our bioinformatics pipeline provides a framework for obtaining such information in the future.

These current limitations notwithstanding, the workflow for identification of BALF microbiota provides a promising framework for integrating gene marker and metaproteomic studies. The development of such bioinformatics pipelines is a critical step toward gaining insight into the distal lung microenvironment and the microbial functional state, as well as the interplay between microbial communities and the host. Improved knowledge about the host-pathogen dynamic may provide the key to unlock more individualized care. The proof-of-concept work described here is a step toward achieving these exciting possibilities. ■

Author disclosures are available with the text of this letter at www.atsjournals.org.

Table 1. Microbial Organisms Identified at the Genus Level with Spectral Identification Statistics, Detection across Clinical Samples in this Study, and Prior Lung Microbiome Studies

Genus	Number of Peptide-Spectrum Matches (Peptides)	Total Number of Datasets*	Total Number of Patients in Whom the Peptide Was Detected	Prior Gene Marker Studies Reporting These Microbiota in the Respiratory Tract
<i>Streptococcus</i>	9 (3) [†]	7	42	(1, 10–12, 14–16)
<i>Parabacteroides</i>	9 (2)	5	30	(10, 11)
<i>Actinomyces</i>	8 (8)	8	44	(1, 10, 11)
<i>Prevotella</i>	8 (6)	6	35	(1, 10, 11, 12, 14, 16)
<i>Haemophilus</i>	6 (4) [†]	1	6	(1, 11, 15)
<i>Cardiobacterium</i>	6 (2)	4	19	(11)
<i>Streptomyces</i>	5 (2)	3	17	
<i>Lactobacillus</i>	4 (4)	4	23	(10, 11)
<i>Enterococcus</i>	3 (3) [†]	3	18	(11)
<i>Anaerofustis</i>	3 (3)	3	18	
<i>Hungatella</i>	3 (3)	3	17	
<i>Paenibacillus</i>	3 (3)	3	16	(10)
<i>Pseudomonas</i>	3 (3)	3	11	(11, 16)
<i>Delftia</i>	3 (2)	2	12	(11)
<i>Pseudoflavonifractor</i>	3 (2)	2	12	
<i>Neisseria</i>	3 (2)	2	11	(1, 10, 11, 14, 16)
<i>Eubacterium</i>	3 (2)	1	6	(11)
<i>Anaerotruncus</i>	3 (1)	1	6	
<i>Pyramidobacter</i>	21 (2)	11	65	
<i>Clostridium</i>	2 (2)	2	12	(10, 11)
<i>Bifidobacterium</i>	2 (2)	2	12	(11, 15)
<i>Ruminococcus</i>	2 (2)	2	12	
<i>Yersinia</i>	2 (2)	1	6	(11)
<i>Succinatimonas</i>	2 (2)	1	6	
<i>Burkholderia</i>	2 (1)	3	18	(11)
<i>Klebsiella</i>	2 (1)	2	12	
<i>Providencia</i>	2 (1)	2	12	
<i>Rothia</i>	2 (1)	1	6	(10–12)
<i>Edwardsiella</i>	2 (1)	1	6	
<i>Fusobacterium</i>	19 (4)	2	12	(1, 10–12, 16)
<i>Bacteroides</i>	14 (11)	5	29	(10)
<i>Oribacterium</i>	12 (4)	4	20	(1, 10, 11)
<i>Capnocytophaga</i>	10 (6)	5	29	(10, 11)
<i>Staphylococcus</i>	1 (1) [†]	1	6	(11, 15, 16)
<i>Scardovia</i>	1 (1)	2	10	(11)
<i>Catonella</i>	1 (1)	1	6	(10, 11)
<i>Treponema</i>	1 (1)	1	6	(10, 11)
<i>Megasphaera</i>	1 (1)	1	6	(10–12, 16)
<i>Anaerococcus</i>	1 (1)	1	6	(11, 15)
<i>Porphyromonas</i>	1 (1)	1	6	(10, 11, 15)
<i>Eikenella</i>	1 (1)	1	6	(11)
<i>Ochrobactrum</i>	1 (1)	1	6	(11)
<i>Alistipes</i>	1 (1)	1	6	(10)
<i>Escherichia</i>	1 (1)	1	6	(10)
<i>Turicibacter</i>	1 (1)	1	6	(10)
<i>Catenibacterium</i>	1 (1)	1	6	
<i>Listeria</i>	1 (1)	1	6	
<i>Paenisporosarcina</i>	1 (1)	1	6	
<i>Turicella</i>	1 (1)	1	6	
<i>Tyzzerella</i>	1 (1)	1	6	
<i>Peptostreptococcus</i>	1 (1)	1	5	(10, 11)
<i>Selenomonas</i>	1 (1)	1	5	(10, 11, 16)
<i>Mycoplasma</i>	1 (1)	1	5	(11)
<i>Shuttleworthia</i>	1 (1)	1	5	(11)
<i>Blautia</i>	1 (1)	1	5	
<i>Collinsella</i>	1 (1)	1	5	
<i>Dorea</i>	1 (1)	1	5	
<i>Kytococcus</i>	1 (1)	1	5	
<i>Oxalobacter</i>	1 (1)	1	5	
<i>Sanguibacter</i>	1 (1)	1	5	

(Continued)

Table 1. (Continued)

Genus	Number of Peptide-Spectrum Matches (Peptides)	Total Number of Datasets*	Total Number of Patients in Whom the Peptide Was Detected	Prior Gene Marker Studies Reporting These Microbiota in the Respiratory Tract
<i>Atopobium</i>	1 (1)	1	4	(10, 11)
<i>Desulfovibrio</i>	1 (1)	1	4	(11)
<i>Parascardovia</i>	1 (1)	1	4	(11)

*Eight-plex runs with one or more peptide spectral matches.

†Genera identified in clinical culture data.

Acknowledgment: The authors thank the Minnesota Supercomputing Institute and the Center for Mass Spectrometry and Proteomics at the University of Minnesota for assistance in this work.

Pratik D. Jagtap, Ph.D.*
University of Minnesota
Minneapolis, Minnesota

Kevin J. Viken, M.B.S.*
University of Minnesota Medical School
Minneapolis, Minnesota

James Johnson
Thomas McGowan, B.S.
University of Minnesota Supercomputing Institute
Minneapolis, Minnesota

Kathryn M. Pendleton, M.D.
University of Minnesota Medical School
Minneapolis, Minnesota

Timothy J. Griffin, Ph.D.
University of Minnesota
Minneapolis, Minnesota

Ryan C. Hunter, Ph.D.
University of Minnesota Medical School
Minneapolis, Minnesota

Joel D. Rudney, Ph.D.
University of Minnesota School of Dentistry
Minneapolis, Minnesota

Maneesh Bhargava, M.D., Ph.D.†
University of Minnesota Medical School
Minneapolis, Minnesota

ORCID IDs: 0000-0003-0984-0973 (P.D.J.); 0000-0003-3822-2269 (K.J.V.); 0000-0003-3248-5738 (K.M.P.); 0000-0001-6801-2559 (T.J.G.); 0000-0003-3841-1676 (R.C.H.); 0000-0002-1294-6181 (M.B.).

*These authors contributed equally to this work.

†Corresponding author (e-mail: bharg005@umn.edu).

References

- Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al.; Lung HIV Microbiome Project. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med* 2013;187:1067–1075.
- Pendleton KM, Erb-Downward JR, Bao Y, Branton WR, Falkowski NR, Newton DW, et al. Rapid pathogen identification in bacterial pneumonia using real-time metagenomics. *Am J Respir Crit Care Med* 2017;196:1610–1612.
- Rebets Y, Lupoli T, Qiao Y, Schirner K, Villet R, Hooper D, et al. Moenomycin resistance mutations in *Staphylococcus aureus* reduce peptidoglycan chain length and cause aberrant cell division. *ACS Chem Biol* 2014;9:459–467.
- Neufeld JD, Dumont MG, Vohra J, Murrell JC. Methodological considerations for the use of stable isotope probing in microbial ecology. *Microb Ecol* 2007;53:435–442.
- Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* 2008;3:e2527.
- Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC II, et al. Community proteomics of a natural microbial biofilm. *Science* 2005;308:1915–1920.
- Bhargava M, Viken KJ, Dey S, Steinbach MS, Wu B, Jagtap PD, et al. Proteome profiling in lung injury after hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant* 2016;22:1383–1390.
- Bhargava M, Viken K, Wang Q, Jagtap P, Bitterman P, Ingbar D, et al. Bronchoalveolar lavage fluid protein expression in acute respiratory distress syndrome provides insights into pathways activated in subjects with different outcomes. *Sci Rep* 2017;7:7464.
- Huson DH, Beier S, Flade I, Górská A, El-Hadidi M, Mitra S, et al. Megan community edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 2016;12:e1004957.
- Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011;184:957–963.
- Hogan DA, Willger SD, Dolben EL, Hampton TH, Stanton BA, Morrison HG, et al. Analysis of lung microbiota in bronchoalveolar lavage, protected brush and sputum samples from subjects with mild-to-moderate cystic fibrosis lung disease. *PLoS One* 2016;11:e0149998.
- Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB, et al. Spatial variation in the healthy human lung microbiome and the adapted island model of lung biogeography. *Ann Am Thorac Soc* 2015;12:821–830.
- Bhargava M, Higgins L, Wendt CH, Ingbar DH. Application of clinical proteomics in acute respiratory distress syndrome. *Clin Transl Med* 2014;3:34.
- Zimmermann A, Zissel G, Müller-Quernheim J, Hofmann S, Schreiber S, Fischer A. Are bronchoalveolar lavages a good source for microbial profiling? Differences between throat and bronchoalveolar lavage microbiomes. *J Med Microbiol* 2015;64:948–951.
- Brown PS, Pope CE, Marsh RL, Qin X, McNamara S, Gibson R, et al. Directly sampling the lung of a young child with cystic fibrosis reveals diverse microbiota. *Ann Am Thorac Soc* 2014;11:1049–1055.

16. Scher JU, Joshua V, Artacho A, Abdollahi-Roodsaz S, Öckinger J, Kullberg S, *et al.* The lung microbiota in early rheumatoid arthritis and autoimmunity. *Microbiome* 2016;4:60.

Copyright © 2018 by the American Thoracic Society

Increased Antielastase Activity in Idiopathic Pulmonary Arterial Hypertension and Chronic Thromboembolic Pulmonary Hypertension

To the Editor:

Pulmonary arterial hypertension (PAH) is characterized by abnormal remodeling and occlusion of precapillary arterioles in the lung with a subsequent increase in pulmonary vascular resistance. This can lead to right ventricular hypertrophy and ultimately right heart failure. Elastase is implicated in the pathobiology of PAH, with evidence including ultrastructural studies showing increased elastase activity in pulmonary arteries from children with congenital heart disease-associated PAH (1), increased elastase release from peripheral blood neutrophils isolated from patients with pulmonary hypertension compared with healthy control subjects (2), and elevated plasma concentrations of elastase in patients with idiopathic PAH (IPAH) (3).

The major elastase inhibitor in the circulation is alpha-1 antitrypsin (AAT). Produced by the liver, AAT is one of the most abundant serine protease inhibitors in the blood, circulating in micromolar concentrations, and accounts for over 90% of neutrophil elastase inhibition in plasma (4–6). AAT protects the lungs against proteolytic damage from elastase, and its deficiency can cause unopposed proteolytic parenchymal damage and emphysema (6).

Increased pulmonary artery elastolytic activity in rat models of PAH can be reversed with elastase inhibitors (7). This suggests that the elastase/AAT axis may be dysregulated in PAH. Although elastase has been measured in plasma from patients with IPAH (3), the levels of endogenous circulating elastase inhibitor AAT remain elusive. One study reported reduced levels of AAT in pooled sera from 20 patients with IPAH based on two-dimensional (2D) gel electrophoresis coupled with mass spectrometry (8). However, 2D gel electrophoresis separates proteins by isoelectric point and size. Plasma AAT is known to exhibit microheterogeneity (9) and is present in multiple isoforms with different isoelectric points due to differential glycosylation (9). Hence, when pooled sera are examined by 2D gel electrophoresis, AAT is represented by multiple spots and the intensity change in one spot is unlikely to reflect the changes in the total amount of AAT.

Supported by British Heart Foundation grants PG/12/54/29734 and PG/15/39/31519 (W.L. and N.W.M.).

Author Contributions: J.G. and K.L. collected and analyzed the data. M.N., K.B., and M.T. collected the patient samples and analyzed the data. N.W.M. holds the ethical approval for collecting the patient plasma samples, has made important contributions to the critical review of the data, and contributed to the drafting of the manuscript. W.L. designed the experiments and collected and analyzed the data. All authors contributed to the writing and critical review of the manuscript.

This letter has a data supplement, which is accessible from this issue's table of contents at www.atsjournals.org.

We measured AAT concentrations in plasma from 29 patients with IPAH, 29 healthy control subjects, and 21 patients with chronic thromboembolic pulmonary hypertension (CTEPH) as a disease comparator (for details regarding the materials and methods used, see the data supplement). The baseline characteristics for the groups are summarized in Table 1. Age and sex did not vary when all three groups were considered ($P = 0.081$ and $P = 0.392$). However, there were individual group differences, with patients with IPAH being younger (median \pm interquartile range: 47 \pm 24 yr) than patients with CTEPH and healthy control subjects, and predominantly female (76%). In contrast to what was reported previously (8), AAT concentrations in plasma assessed by ELISA did not vary significantly between patients with IPAH (mean \pm SEM: 1.91 \pm 0.04 g/L) and healthy control subjects (1.79 \pm 0.05 g/L; $P = 0.052$; Figure 1A). There was also no difference between healthy control subjects and patients with CTEPH (1.81 \pm 0.09 g/L; $P = 0.775$; Figure 1A). We found slightly higher elastase-inhibitory activities in plasma from patients with IPAH and patients with CTEPH (Figure 1B). The ELISA did not show a similar increase in AAT levels, possibly because elastase inhibitors other than AAT were present, or the heterogeneity of the AAT glycosylation resulted in some isoforms being more reactive with the antibodies used in the ELISA measurement.

We next examined plasma AAT using SDS-PAGE and immunoblotting. AAT is a member of the serine protease inhibitor (SERPIN) family. Its native form has a long reactive center loop (RCL) acting as the bait; hence, SERPINs are suicidal protease inhibitors (10). Upon encountering a target protease, the RCL is recognized and bound by the protease, allowing the formation of a Michaelis complex (Figure 1C). A covalent bond is then formed between the RCL and the protease active-site residue, followed by cleavage of the RCL and insertion of the RCL into β -sheet A, which brings the covalently linked protease to the opposing end of the SERPIN molecule (10). During this process, cleaved AAT can be generated either by protease cleavage before the final complex formation (Figure 1C, arrow 1) or by breakdown of the final complex (Figure 1C, arrow 2) (10), both of which occur as a result of protease activity. The resulting cleaved AAT contains two peptide fragments (Figure 1C, cleaved AAT, green and blue), which can be separated by SDS-PAGE. We used an antihuman AAT antibody that can detect both native AAT and the larger fragment of the cleaved AAT, which appeared as a distinct lower band owing to its lower molecular weight. As shown in Figure 1D, AAT in plasma from both patients with IPAH and healthy control subjects is predominantly in the native form (upper band), which can be converted into the cleaved form (lower band) upon incubation with elastase, in a manner identical to that used for recombinant AAT. When equal volumes of plasma from healthy control subjects, patients with IPAH, and patients with CTEPH were run simultaneously on SDS-PAGE and with longer exposure, more cleaved AAT was detected in plasma from both patients with IPAH and patients with CTEPH (Figure 1E and F), suggesting more protease activity in the plasma from these patients. Interestingly, the two major cleaved bands C1 and C2 (Figure 1E) are smaller than the major elastase cleavage product C3, suggesting there are proteases in plasma other than elastase that could cleave AAT. It is well known that activated neutrophils release two other serine proteases, proteinase 3 and cathepsin G, both of which can be inhibited by AAT and produce cleaved AAT.