bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Original Research

Title: COVID-19 lung disease shares driver AT2 cytopathic features with Idiopathic pulmonary fibrosis

Authors: Saptarshi Sinha[3]*, Vanessa Castillo[3], Celia R. Espinoza[3], Courtney Tindle[3], Ayden G. Fonseca[3], Jennifer M. Dan[4, 5], Gajanan D. Katkar[3], Soumita Das[6] Debashis Sahoo[1, 2]†* and Pradipta Ghosh[3, 7]†

Affiliations:

[1]*Department of Pediatrics, University of California San Diego, La Jolla, CA, 92093.*
[2]*Department of Computer Science and Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, CA, 92093.*
[3]*Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, 92093.*
[4]*Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology (LJI), La Jolla, CA, USA.*
[5]*Department of Medicine, Division of Infectious Diseases and Global Public Health, University of California, San Diego (UCSD), La Jolla, CA, USA.*
[6]*Department of Pathology, University of California San Diego, La Jolla, CA, 92093.*
[7]*Department of Medicine, University of California San Diego, La Jolla, CA, 92093.*

*Equal contribution
† Co-Corresponding

Corresponding authors:

Debashis Sahoo, Ph.D.; Associate Professor, Department of Pediatrics, University of California San Diego; 9500 Gilman Drive, MC 0703, Leichtag Building 132; La Jolla, CA 92093-0703. **Phone:** 858-246-1803: **Fax:** 858-246-0019: **Email:** dsahoo@ucsd.edu

Pradipta Ghosh, M.D.; Professor, Departments of Medicine and Cell and Molecular Medicine, University of California San Diego; 9500 Gilman Drive (MC 0651), George E. Palade Bldg, Rm 232; La Jolla, CA 92093. **Phone**: 858-822-7633:
Email: prghosh@ucsd.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

STRUCTURED ABSTRACT

Background: In the aftermath of Covid-19, a long-haul form of mysterious and progressive fibrotic lung disease has emerged, i.e., *p*ost-*C*OVID-19 *l*ung *d*isease (PCLD), for which we currently lack insights into pathogenesis, disease models, or treatment options.

Method: Using an AI-guided approach, we analyzed > 1000 human lung transcriptomic datasets associated with various lung conditions using two viral pandemic (ViP and sViP) and one covid lung gene signatures. Upon identifying similarities between COVID-19 and idiopathic pulmonary fibrosis (IPF), we subsequently dissected the basis for such similarity from molecular, cytopathic, and immunologic perspectives using a panel of IPF-specific gene signatures, alongside signatures of alveolar type II (AT2) cytopathies and of prognostic monocyte-driven processes that are known drivers of IPF. To pinpoint the AT2 processes that are shared points of convergence between COVID-19 and IPF, transcriptome-derived findings were used to construct protein – protein interaction (PPI) network. Key findings were validated in hamster and human adult lung organoid (ALO) pre-clinical models of COVID-19 using immunohistochemistry and qPCR.

Findings: We found that COVID-19 resembles IPF at a fundamental level; it recapitulates the gene expression patterns (ViP and IPF signatures), cytokine storm (IL15-centric) and the AT2 cytopathic changes, e.g., injury, DNA damage, arrest in a transient, damage-induced progenitor state, and senescence-associated secretory phenotype (SASP). These immunocytopathic features were induced in pre-clinical COVID models (ALO and hamster) and reversed with effective anti-CoV-2 therapeutics in hamsters. PPI-network analyses pinpointed ER stress as one of the shared early triggers of both diseases, and IHC studies validated the same in the lungs of deceased subjects with COVID-19 and SARS-CoV-2-challenged hamster lungs. Lungs from *tg*-mice, in which ER stress is induced specifically in the AT2 cells, faithfully recapitulate the host immune response and alveolar cytopathic changes that are induced by SARS-CoV-2.

Interpretation: Like IPF, COVID-19 may be driven by injury-induced ER stress that culminates into progenitor state arrest and SASP in AT2 cells. The ViP signatures in monocytes may be key determinants of prognosis. The insights, signatures, disease models identified here are likely to spur the development of therapies for patients with IPF and other fibrotic interstitial lung disease.

Funding: This work was supported by the National Institutes for Health grants R01- GM138385 and AI155696 and funding from the Tobacco-Related disease Research Program (R01RG3780).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

One Sentence Summary: Severe COVID-19 triggers cellular processes seen in fibrosing Interstitial Lung Disease

Key Words:

- Artificial Intelligence/Machine Learning
- Boolean Equivalent Clusters
- Coronavirus COVID-19
- COVID-19 lung
- Immune response
- Alveolar Type II Pneumocytes
- Damage-associated transient progenitor (DATP)
- Senescence
- ER-stress
- Interleukin 15 (IL15)

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

PANEL: RESEARCH IN CONTEXT

Evidence before this study

In its aftermath, the COVID-19 pandemic has left many survivors, almost a third of those who recovered, with a mysterious long-haul form of the disease which culminates in a fibrotic form of interstitial lung disease (post-COVID-19 ILD). Post-COVID-19 ILD remains a largely unknown entity. Currently we lack insights into the core cytopathic features that drives this condition.

Added value of this study

Using an AI-guided approach, which involves the use of a sets of gene signatures, protein-protein network analysis, and a hamster model of COVID-19, we have revealed here that COVID-19 -lung fibrosis resembles IPF, the most common form of ILD, at a fundamental level—showing similar gene expression patterns in the lungs and blood, and dysfunctional AT2 processes (ER stress, telomere instability, progenitor cell arrest and senescence). These findings are insightful because AT2 cells are known to contain an elegant quality control network to respond to intrinsic or extrinsic stress; a failure of such quality control results in diverse cellular phenotypes, of which ER stress appears to be a point of convergence, which appears to be sufficient to drive downstream fibrotic remodeling in the lung.

Implications of all the available evidence

Because unbiased computational methods identified the shared fundamental aspects of gene expression and cellular processes between COVID-19 and IPF, the impact of our findings is likely to go beyond COVID-19 or any viral pandemic. The insights, tools (disease models, gene signatures, and biomarkers), and mechanisms identified here are likely to spur the development of therapies for patients with IPF and other fibrotic interstitial lung disease, all of whom have limited or no treatment options. to dissect the validate prognostic biomarkers to assess and track the risk of pulmonary fibrosis and develop therapeutics to halt fibrogenic progression.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

As the acute phase of the COVID-19 pandemic winds down, the chronic diseases in its aftermath have begun to emerge. For example, many survivors are suffering from a mysterious long-haul form of the disease which culminates in a fibrotic form of interstitial lung disease (post-COVID-19 ILD)[1-9]. The actual prevalence of post-COVID-19 ILD (henceforth, *PCLD*) is still emerging; early analysis indicates that more than a third of the survivors develop fibrotic abnormalities. One of the important determinants for PCLD[10] is the duration of disease; ~4% of patients with a disease duration of less than 1 week, ~24% of patients with a disease duration between 1-3 weeks, and ~61% of patients with a disease duration > 3 weeks, developed fibrosis. Disability among survivors of severe COVID-19 has been primarily attributed to reduced lung capacities[11]. While COVID-19–affected lungs and post-Influenza lungs share the histologic pattern of diffuse alveolar damage (DAD)[12] with perivascular T-cell infiltration, key differences have been observed[13]. The lungs from patients with Covid-19 also showed[13]: (i) widespread thrombosis with microangiopathy; (ii) severe endothelial injury associated with the presence of intracellular virus and disrupted cell membranes; and (iii) neovascularization predominantly through a mechanism of intussusceptive angiogenesis. These findings highlight that despite being viral pandemics, COVID-19, but not influenza impacts the lungs in ways that go beyond acute DAD, a common phenomenon in progressive ILDs[12]. Although a recent study[14] implicated pathogenic subsets of respiratory $CD8^+$ T cells in contributing to persistent tissue microenvironment in PCLD, what remains largely unknown are some of the earliest cellular and molecular mechanisms that fuel the progression of fibrosis in COVID-19 lung (but not influenza).

As for how the fibrotic sequel of COVID-19 is managed at present, based on the shared pathological features of overt fibrosis in both end-stage COVID-19 and ILDs, lung transplantation remains the mainstay option for these patients[15-17]. Although corticosteroids appear to improve the risk of PCLD[2], beyond that, there is no existing therapeutic option.

Here we seek to unravel the fundamental molecular mechanisms underlying PCLD, identifying key disease drivers (cellular processes, immune pathways, and the signaling cascades that support those pathways). We use artificial intelligence (AI) and machine learning derived gene signatures, the *Vi*ral *P*andemic, ViP and severe(s) ViP signatures that are induced in all respiratory viral pandemics[18]. Besides the ViP signatures, we also use gene signatures induced in the lungs of patients with severe COVID-19. These approaches, in conjunction with experimental validation (in pre-clinical disease models and in human lung tissue) not only helped identify which lung pathology shares fundamental molecular features with COVID-19 lung, but also revealed key mechanistic insights into the pathogenesis of PCLD. The findings also provide clues into how to navigate, prognosticate and recapitulate in pre-clinical models this emergent mysterious condition. The objectivity and precision of the AI-guided unbiased approaches enhance the translational potential of our findings.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

METHODS

KEY RESOURCE TABLE:

MATERIALS & REAGENTS			
ANTIBODIES USED FOR IMMUNOCYTOCHEMISTRY			
Name	**Manufacturer**	**Catalog number**	**Dilution factor**
p14 ARF (E3X6D)	Cell Signaling	74560S	1:250
GRP78/BIP Rabbit polyclonal antibody	Proteintech	11587-1-AP	1:500
P53 Rabbit polyclonal antibody	Proteintech	21891-1-AP	1:200
Cytokeratin 8 Rabbit polyclonal antibody	Proteintech	10384-1-AP	1:2000
Claudin 4-specific Rabbit polyclonal antibody	Proteintech	16195-1-AP	1:200
p21 Waf1 / Cip1 (12D1)	Cell Signaling	2947S	1:50
INSTRUMENTS			
Leica DMI4000B (Automated Inverted Microscope)	Leica Microsystems	DMI4000B	
Power Pressure Cooker XL	Tristar Products		
MESO QuickPlex SQ 120	PBL Assay Science		
SOFTWARE & PACKAGES			
ImageJ	https://imagej.nih.gov/ij/index.html		
GraphPad Prism	https://www.graphpad.com/scientific-software/prism/		
MSD® DISCOVERY WORKBENCH 4.0	https://www.mesoscale.com/en/products_and_services/software/		
NetworkX	https://networkx.org		
Cytoscape	https://cytoscape.org		
KITS, ENZYMES, CHEMICALS, AND REAGENTS			
Zinc Formalin	Fisher Scientific	23-313096	
Xylene	VWR	XX0060-4	
Hematoxylin	BioGenex	HK100-9K	
Two-Component DAB	BioGenex	HK542-XAKE	2 drops per 1mL
Ethanol	Koptec	UN1170	
Sodium Citrate	Sigma-Aldrich	W302600	
Tris Base	Fisher Scientific	BP154-1	
Ethylenediaminetetraacetic acid (EDTA)	Fischer Scientific	BP118-500	
3% Hydrogen Peroxide	Target	245-07-3628	
ImmPRESS HRP Horse Anti-Rabbit IgG Polymer Detection Kit, Peroxidase	Vector Labs	MP-7401	
Goat serum 2.5%	Vector Laboratories	S-10120	
Normal Antibody Diluent	MP Biomedicals, LLC	980641	

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DETAILED METHODS:

Computational Methods

ViP and severe (s)ViP Signatures

ViP (Viral Pandemic) signature[18] is derived from a list of 166 genes using Boolean Analysis of large viral infection datasets (training datasets: GSE47963, n = 438; GSE113211, n = 118). This 166-gene signature was conserved in all viral pandemics, including COVID-19, inspiring the nomenclatures ViP signature. A subset of 20-genes classified disease severity called severe-ViP signature using an additional cohort (GSE101702, n = 159)[18]. No interstitial lung disease (ILD) datasets, or for that matter any other lung dataset was used during the process of discovery and validation of the ViP/sViP signatures. To compute the ViP signature, first the genes present in this list were normalized according to a modified Z-score approach centered around StepMiner threshold (formula = (expr -SThr)/3*stddev). The normalized expression values for every probeset for 166 genes were added together to create the final ViP signature. The severe ViP signature is computed similarly using 20 genes. The samples were ordered finally based on both the ViP and severe-ViP signature. A color-coded bar plot is combined with a violin plot to visualize the gene signature-based classification.

Protein-protein interaction network (PPIN) construction

PPIN has been constructed using the gene signatures as a set of seed nodes. The nodes between the seed nodes were fetched using the connecting shortest paths and their components from the human protein interaction dataset of the STRING database[19]. A high cutoff of STRING interaction score has been chosen based on the proteins present in the signature list to neglect the false positive interactions.

Data analysis

Several publicly available microarrays and RNASeq databases were downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) website.[20-22] Gene expression summarization was performed by normalizing Affymetrix platforms by RMA (Robust Multichip Average)[81, 82] and RNASeq platforms by computing TPM (Transcripts Per Millions) [23, 24] values whenever normalized data were not available in GEO. We used log2(TPM +1) as the final gene expression value for analyses. GEO accession numbers are reported in figures, and text. A catalog of all datasets analyzed in this work can be found in **Supplementary Table 1.** StepMiner analysis and methodologies for composite gene signatures, outcome analyses, and violin plots are detailed in Supplementary Online Methods.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Statistical Analyses

Gene signature is used to classify sample categories and the performance of the multi- class classification is measured by ROC-AUC (Receiver Operating Characteristics Area Under the Curve) values. A color-coded bar plot is combined with a density plot to visualize the gene signature-based classification.All statistical tests were performed using R version 3.2.3 (2015-12-10). Standard t-tests were performed using python scipy.stats.ttest_ind package (version 0.19.0) with Welch's Two Sample t-test (unpaired, unequalvariance (equal_var=False), and unequal sample size) parameters. Multiple hypothesis correction was performed by adjusting *p* values with statsmodels.stats.multitest.multipletests (fdr_bh: Benjamini/Hochberg principles). The results were independently validated with R statistical software (R version 3.6.1; 2019-07-05). Pathway analysis of gene lists were carried out via the Reactome database and CluGo algorithm. Reactome identifiessignaling and metabolic molecules and organizes their relations into biological pathways and processes. Kaplan- Meier analysis is performed using lifelines python package version 0.14.6. Violin, Swarm and Bubble plots are created using python seaborn package version 0.10.1. The source code for Boolean analysis framework is available at https://github.com/sahoo00/BoNE[25] and https://github.com/sahoo00/Hegemon[26].

Experimental Methods

Rapid autopsy procedure for tissue collection

The lung specimens from the COVID-19 positive human subjects were collected as described in detail previously [18, 27] using autopsy (study was IRB Exempt). All donations to this trial were obtained after telephone consent followed by written email confirmation with next of kin/power of attorney per California state law (no in-person visitation could be allowed into our COVID-19 ICU during the pandemic). The team member followed the CDC guidelines for COVID-19 and the autopsy procedures[8, 9]. Lung specimens were collected in 10% Zinc-formalin and stored for 72 h before processing for histology. Autopsy #2 was a standard autopsy performed by anatomical pathology in the BSL3 autopsy suite. The patient expired and his family consented for autopsy. After 48 hours, lungs were removed and immersion fixed whole in 10% formalin for 48 hours and then processed further. Lungs were only partially fixed at this time (about 50% fixed in thicker segments) and were sectioned further into small 2-4cm chunks and immersed in 10% formalin for further investigation. Autopsies #4 and #5 were collected from rapid postmortem lung biopsies. The procedure was performed in the Jacobs Medical Center ICU (all the ICU rooms have a pressure-negative environment, with air exhausted through HEPA filters [Biosafety Level 3 (BSL3)] for isolation of SARS-CoV-2 virus). Biopsies were performed 2-4 hours after patient expiration. Ventilator was shut off to reduce aerosolization of viral particles at least 1 hour after loss of pulse and before the sample collection. Every team member had personal protective equipment in accordance with the University policies for procedures on patients with COVID-19 (N95 mask + surgical mask, hairnet, full face shield, surgical gowns,

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

double surgical gloves, booties). Lung biopsies were obtained after L-thoracotomy in the 5th intercostal space by our cardiothoracic surgery team. Samples were taken from the left upper lobe (LUL) and left lower lobe (LLL) and then sectioned further.

Collection of blood from COVID-19 patients

Blood from COVID-19 donors was either obtained at a UC San Diego Health clinic under the approved IRB protocols of the University of California, San Diego (UCSD; 200236X) or recruited at the La Jolla Institute under IRB approved (LJI; VD-214). COVID-19 donors were California residents, who were either referred to the study by a health care provider or self-referred. Blood was collected in acid citrate dextrose (ACD) tubes (UCSD) or in EDTA tubes (LJI) and stored at room temperature prior to processing for plasma collection. Seropositivity against SARS-CoV-2 was confirmed by ELISA. At the time of enrollment, all COVID-19 donors provided written informed consent to participate in the present and future studies. Patient characteristic is listed in Supplemental Information 3.

Human serum cytokines measurement

Human serum cytokines measurement was performed using the V-PLEX Custom Human Biomarkers from MSD platform [MESO QuickPlex SQ 120]. Human serum samples fractionated from peripheral blood of KD and MIS-C patients (all samples collected prior to the initiation of treatments) were analyzed using customized standard multiplex plates as per the manufacturer's instructions using recommended software [MSD® DISCOVERY WORKBENCH 4.0].

COVID-19 lung models in-a-dish

Monolayers derived from adult lung organoids (ALOs), primary airway cells, or hiPSC-derived alveolar type II (AT2) pneumocytes were infected with SARS-CoV-2 to create in vitro lung models of COVID-19 as described previously[27]. The raw data and processed data were deposited in Gene Expression Omnibus under accession no. GSE157057. These datasets were analyzed here using various gene signatures.

COVID-19 modeling in Syrian hamsters

Lung samples from 8-week-old Syrian hamsters were generated from experiments conducted exactly as in a previously published study[28]. Animal studies were approved and performed in accordance with Scripps Research IACUC Protocol #20-0003[28]. We chose three different groups of samples: uninfected control, SARS-CoV-2 challenge after Den3 (antibody to dengue virus), and SARS-CoV-2 challenge after Anti-CoV2 (CC12.2; a potent SARS-CoV-2 neutralizing antibodies)[28].

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Immunohistochemistry

COVID-19 samples were inactivated by storing in 10% formalin for 2 days and then transferred to zinc-formalin solution for another 3 days. The deactivated tissues were transferred to 70% ethanol and cassettes were prepared for tissue sectioning. The slides containing hamster and human lung tissue sections were deparaffinized in xylene (Sigma-Aldrich Inc., MO, USA; catalog# 534056) and rehydrated in graded alcohols to water. For GRP78/ BIP, p53, Cytokeratin 8, and Claudin 4-specific polyclonal antigen retrieval, slides were immersed in Tris-EDTA buffer (pH 9.0) and boiled for 10 minutes at 100°C. Slides were immersed in Sodium Citrate Buffer (pH 6.0) and boiled for 10 minutes at 100°C, for p14 ARF (E3X6D) and p21 Waf1/ Cip1(12D1) antigen retrieval. Endogenous peroxidase activity was blocked by incubation with 3% H2O2 for 5-10 minutes. To block non-specific protein binding either 2.5% goat or 2.5% horse serum (Vector Laboratories, Burlingame, USA; catalog# MP-7401 or S-1012) was added. Tissues were then incubated with the following antibodies: rabbit GRP78/ BIP polyclonal antibody (1:500 dilution; proteintech®, Rosemont, IL, USA; catalog# 11587-1-AP), rabbit p53 polyclonal antibody (1:200 dilution ; proteintech®, Rosemont, IL, USA; catalog# 21891-1-AP), rabbit Cytokeratin 8 polyclonal antibody (1:2000 dilution; proteintech®, Rosemont, IL, USA; catalog# 10384-1-AP), and rabbit Claudin 4-specific polyclonal antibody (1:200 dilution, proteintech®, Rosemont, IL, USA; catalog# 16195-1-AP) for 1.5 hours at room temperature in a humidified chamber then rinsed with TBS twice for 3 minutes each. Antibodies were prepared in antibody diluent (MP Biomedicals, LLC., Catalog # 980641) for rabbit p14 ARF (E3X6D) monoclonal antibody (1:250 dilution; Cell Signaling Technology, Danvers, MA, USA; catalog# 74560S) and rabbit p21 Waf1/ Cip1 (12D1) monoclonal antibody (1:50 dilution; Cell Signaling Technology, Danvers, MA, USA; catalog# 2947S) that were then incubated in a humidified chamber overnight at 4°C, then rinsed with TBST twice for 5 minutes each. Sections were incubated with horse anti-rabbit (Vector Laboratories, Burlingame, USA; catalog# MP-7401) secondary antibodies for 30 minutes at room temperature in a humidified chamber and then washed with TBS or TBST 3x, 5 minutes each. Then incubated with DAB (BioGenex, Fremont, CA, USA; catalog # HK542-XAKE) for 5-10 minutes and counterstained with hematoxylin (BioGenex, Fremont, CA, USA; catalog# HK100-9K), dehydrated in graded alcohols, cleared in xylene, and cover slipped. All sides were visualized by Leica DM1000 LED (Leica Microsystems, Germany).

IHC Quantification

IHC images were randomly sampled at different 300x300 pixel regions of interest (ROI). The ROIs were analyzed using IHC Profiler[29]. IHC Profiler uses a spectral deconvolution method of DAB/hematoxylin color spectra by using optimized optical density vectors of the color deconvolution plugin for proper separation of the DAB color spectra. The histogram of the DAB intensity was divided into 4 zones: high positive (0 to 60), positive (61 to 120), low positive (121 to 180) and negative (181 to 235). High positive, positive, and low positive percentages

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were combined to compute the final percentage positive for each region of interest (ROI). The range of values for the percent positive is compared among different experimental groups.

Data availability

Source data are provided with this paper. All data is available in the main text or the supplementary materials. Publicly available datasets used are enlisted in **Supplemental Information 1**.

Code availability

The software codes are publicly available at the following links: https://github.com/sahoo00/BoNE[25] and https://github.com/sahoo00/Hegemon 86[26]. The PPI network analysis can be found at: https://github.com/sinha7290/PPIN.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESULTS

A study design that uses gene signatures as a computational framework to navigate COVID-19 lung disease

Because PCLD is still an emergent illness that lacks insights into disease pathophysiology, we resorted to a study design (**Figure 1**) that is geared to achieve three goals: (i) maximize rigor by using diverse transcriptomic datasets cohorts (a total of 1556 unique samples, human: 1497; mouse: 41, hamster: 18; see **Supplemental Information 1**), but (ii) minimize bias by using well characterized sets of gene signatures from independent publications (see **Supplemental Information 2**), and (iii) retain objectivity and precision by using artificial intelligence/machine learning (AI/ML) tools that are supported by precise mathematical algorithms and statistical tools (see *Methods*).

First, to ensure that the findings maintain relevance to respiratory viral pandemics and COVID-19, we used a 166-gene signature that is conserved in all *vi*ral *p*andemics (ViP), including COVID-19, and a subset of 20-genes within that signature that classifies disease severity[18] (see **Supplemental Information 2**). In the absence of a sufficiently large number of COVID-19 datasets at the onset of the COVID-19 pandemic, these ViP signatures were validated on ~45000 datasets representing the pandemics of the past (Training datasets: Influenza and avian flu; GSE47963, n = 438; GSE113211, n = 118; Validated in numerous datasets as cited before) and used without further training to prospectively analyze the samples from the current pandemic (i.e., COVID-19) (**Figure 1**, *Step 0*). The ViP signatures appeared to capture the 'invariant' host response, i.e., the shared fundamental nature of the host immune response induced by all viral pandemics, including COVID-19. Since they were discovered, these signatures have helped navigate the syndrome of multisystem inflammatory syndrome in children[30] (MIS-C; a.k.a PIMS-TS)[31]. MIS-C is still a relatively poorly understood entity, but none-the-less recognized as an immunologic response to viral exposure that shares many clinical features with the syndrome of Kawasaki disease (KD). The ViP signatures helped revealed that KD and MIS-C are on the same continuum of the host immune response as COVID-19. Here, we used the ViP signatures as a starting computational framework to navigate the syndrome of COVID-19 lung disease.

We begin first by using the ViP signatures in conjunction with an independently reported[13] COVID-19-lung derived signature to sift through diverse lung diseases and identify what pathologic condition comes closest to COVID-19 lung disease; this effort led us to pinpoint the prototypical example of interstitial lung disease that is characterized by progressive fibrosis, i.e., idiopathic pulmonary fibrosis (IPF)[32] as a computational match for COVID-19 lung disease (**Figure 1**, *Step 1*). The subsequent steps in the study use the ViP signatures alongside IPF-specific PBMC-derived prognostic signatures (**Figure 1**, *Step 2*) and signatures of alveolar cytopathic changes (**Figure 1**, *Step 3*) to systematically assess the degree of similarities between COVID-19 lung and IPF and the crossover of pathophysiologic processes in the two conditions. In the final step (**Figure 1**, *Step 4*), a set of gene signature-inspired protein-protein interaction (PPI) network analysis is used which pinpointed alveolar

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ER stress as a potential early step in both IPF and PCLD, which is sufficient to recapitulate not just the ViP signatures, but also the alveolar cytopathic features that are shared between the diseases. Key predictions are validated in human lung tissues and plasma from COVID-19 subjects, hamster models of COVID-19 and in human pre-clinical lung models (**Figure 1**, *Steps 3-4*).

COVID-19 lung disease and IPF induce a common set of gene expression signatures

We began by asking if the ViP/sViP signatures that were found to be induced in all CoV samples tested so far[18], are induced in certain lung diseases and not in others. To avoid the overreliance on the ViP signatures, we included a second set of CoV-lung disease gene signatures which were derived from a differential expression analysis on lung samples from healthy controls vs. fatal COVID-19[13] (see **Supplemental Information 2**). As for the lung diseases, we ensured that all four major pathologic conditions were represented: neoplastic (e.g., carcinoids and adenocarcinoma), granulomatous (e.g., tuberculosis/TB and sarcoidosis), allergic/infectious (e.g., severe asthma and community acquired pneumonia/CAP and tuberculosis) and vasculopathic (e.g., pulmonary hypertension/PHT and chronic thromboembolic pulmonary hypertension/CTEPH) (**Figure 2A**). As expected, ViP/sViP signatures were induced in infectious diseases (CAP and tuberculosis), and to our surprise, these signatures were upregulated also in sarcoidosis, a granulomatous condition that progresses to fibrosis in ~20% of the patients[33, 34] (**Figure 2A**). Only two diseases (tuberculosis and IPF) showed significant induction of both ViP and CoV-lung signatures, of which only one condition (IPF) manifests as a diffuse disease with patchy parenchymal involvement that has been reported in COVID-19 lung disease (**Figure 2A**).

We next analyzed a set of six signatures, each independently derived from diverse IPF-centric studies[35-38], on diverse CoV samples; we also analyzed the ViP/sViP[18] and CoV-lung[13] signatures on 3 independent IPF datasets. We found that most CoV samples induced the IPF signatures and *vice versa*, i.e., most IPF samples induced ViP/CoV-lung signatures (**Figure 2B**). In fact, in a pooled dataset that included both CoV and IPF samples the ViP (**Figure 2C**), sViP (**Figure 2D**), IPF-lung (**Figure 2E**) and CoV-lung (**Figure 2F**) were indistinguishably induced in both samples compared to healthy controls. These results indicate that CoV and IPF signatures have crossover utility for sample classification and indicate that IPF is the closest computational match to COVID-19 lung disease (**Figure 2G**). Consistent with the fact that PCLD occurs in survivors of severe disease that required hospitalization[39], we found that sViP signature was induced significantly higher in hospitalized compared to non-hospitalized subjects (**Figure 2H**). Finally, we asked what, if any relationship exists between another poorly understood feature of long-haul COVID-19, i.e., severe fatigue[40, 41]. To improve rigor, we analyzed a set of other well-defined fatigue syndromes (e.g., gulf war illness, chronic fatigue syndrome, and idiopathic chronic fatigue) and for consistency, restricted the analysis to only whole blood/PBMC samples (**Figure 2I**). ViP/sViP and CoV-lung immune signatures were induced only in CoV samples, but not in the remaining fatigue-

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

associated syndromes (**Figure 2I**), indicating that gene expression signatures in long-haul COVID-19 are distinct from other fatigue-associated syndromes. Instead, fatigue in long-haul COVID-19 may be of similar reasons/pathophysiology that has been observed in IPF [42]. It is noteworthy that profound fatigue is a shared phenomenon in both IPF and sarcoidosis[42, 43], both carry the risk of progressive fibrosis, and both conditions induce the ViP/sViP response (**Figure 2A**).

Taken together, these results show that when it comes to gene expression patterns, long-haul PCLD is closest to IPF: both are characterized by profound fatigue, diffuse and progressive lung fibrosis, and share gene expression patterns. The latter suggests that these two diseases may also share molecular and cellular mechanisms of disease pathogenesis.

PBMC-derived prognostic signatures in COVID-19 and IPF show crossover utility in both diseases. Previously we showed using single cell sequencing datasets in COVID-19, as well as datasets in which bulk seq was performed on sorted cells, that ViP/sViP signatures are primarily induced by lung epithelial cells and myeloid cells (both alveolar macrophages and PBMCs)[18]. When assessed in whole blood and/or PBMCs, the sViP signature was prognostic in COVID-19—high expression of this 20-gene signature was associated with prolonged hospital stay and mechanical ventilation[18]. We noted that a PBMC-based 52-gene signature has previously been validated in numerous IPF datasets[44]. In the setting of IPF, the role of macrophages has been cemented by several studies. Derived either from tissue-residentmacrophages or PBMCs, macrophages are first activated by the secreted cytokines from senescent alveolar epithelial cells. Subsequently, they go on to activate the alveolar fibroblasts via secretionof CCL18, TGFβ, IL-1β, and/or PDGF, promoting their secretion of collagen[45-50] (**Figure 3A**). In later stages of IPF, the activated macrophages and myofibroblasts are thought toeven cross-stimulate each other,resulting in a vicious cycle that assures propagation of fibrosisthroughout the lung[45, 51].

Consistent with what others have reported before[23, 52], we found that alveolar macrophages (marked by CD-14[53]) in the lungs of deceased subjects contained the viral nucleocapsid protein (**Figure 3B-C**). The 20-gene sViP signature was prognostic in the blood datasets from both COVID-19 (**Figure 3D**; *left*) and IPF subjects (**Figure 3D**; *middle, right*); among IPF subjects, the signature was prognostic exclusively in male subjects (**Figure 3D**; *middle*). The 52-gene 'IPF-specific' prognostic signature[47] retained its ability to prognosticate outcome in both male and female patients with IPF (as expected; **Figure 3E**; *middle, right*), but also did so in the setting of COVID-19 (**Figure 3E**; *left*). By contrast, the CoV-lung tissue-derived signature that is known to distinguish CoV from both healthy and influenza-affected lungs[54], did not prognosticate outcome in either COVID-19 (**Figure 3F**; *left*) or IPF subjects (**Figure 3F**; *middle vs. right*). It indicates that the PBMC-derived signatures in CoV and IPF retain their prognostic relevance in both conditions, with at least one key notable difference; sViP, but not the IPF signature revealed the well-established gender specific differences in the progression of IPF.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Taken together, these results indicate that PBMCs in both IPF and COVID-19 have similar gene induction patterns, and that these circulating immune cells may be determinants of outcome in both conditions. Our findings are in keeping with the fact that a subset of pro-fibrogenic macrophages have been detected in end-stage severe COVID-19 that progressed to fibrosis requiring lung transplantation[15]. Others have reported that these profibrogenic macrophages express markers associated to fibrotic processes (like *TREM2* or *SPP1*) in patients with severe COVID-19[55]; *TREM2* is suggested to prevent macrophage apoptosis and promote chronic inflammatory disease after lung viral infection[56]. We conclude that the prognostic value of IPF- and sViP signatures induced in the PBMCs may reflect the profibrogenic immune response that is shared in both conditions.

Alveolar type II (AT2) telomere dysfunction and senescence are shared features in both COVID-19 and IPF.

Besides the contributions from PBMCs, a contemporary paradigm in IPF is that chronic injury to distal lung tissue leads to either loss or altered function of epithelial stem cells (i.e., AT2 cells), that promote dysregulated repair and pathogenic activation of fibroblasts.[57-59] We asked if the AT2 dysfunction states previously reported in IPF also occur in COVID-19. To this end, we analyzed a single cell dataset of bronchoalveolar lavage from COVID-19 patients (GSE145926) for the induction of ViP/sViP, CoV-lung and IPF signatures and other signatures that are indicative of three important AT2 cytopathies: (i) Damage associated transient progenitor (DATP[60]), a distinct *KRT8*+/*CLDN4*+ AT2 lineage that has been independently shown, nearly simultaneously by 3 independent groups[6061, 62] to be associated with activated myofibroblasts during lung fibrosis in the setting of IPF[28]. (ii) AT2-senescence signature, which was derived from another IPF-focused study[63] that used human and mouse models to pinpoint Doxycycline-induced program of p53-dependent cellular senescence, AT2 cell depletion, and spontaneous, progressive pulmonary fibrosis; and (iii) Aging associated telomerase dysfunction (TERC_UP) signature, which was derived from aging telomerase knockout (*Terc*$^{-/-}$) mice. We found that the entire panel of signatures analyzed were upregulated, most consistently, only in the epithelial cells (**Figure 4A**), whereas only some were induced and not others in the non-epithelial compartments.

We next confirmed that a composite score of *IL15/IL15RA*, AT2-senescence and telomerase dysfunction (TERC_UP) signatures were induced indistinguishably in both CoV and IPF samples compared to healthy controls (**Figure 4B**). A correlation matrix (**Figure 4C**) on the entire cohort showed that while both IL15 and IL15RA correlate with the ViP signature (as expected), IL15RA, but not IL15 correlate with TERC_UP and AT2-senescence signatures. Both ViP and sViP signatures correlate with AT2 senescence. Furthermore, when we tested AT2-senescence and telomere dysfunction in other CoV datasets, we found that these features were conserved in most CoV samples across diverse cohorts (**Figure 4D**). These findings are in keeping with prior IPF modeling efforts using bleomycin-challenged transgenic mice that lack the *Telomeric Repeat Binding Factor 2* in their AT2 cells (*Trf2*$^{Fl/Fl}$*; Sftpc-CreER*) have shown that senescent AT2 cells with telomere dysfunction

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

upregulate immune-signaling pathways that is IL15-centric[64], which is the central cytokine pathway in ViP signature and COVID-19[18].

Taken together, these findings support the existence of telomere dysfunction and DNA damage in both acute COVID-19 and IPF. Mutations in the two main components of the telomerase holoenzyme complex that is responsible for creating new telomeric DNA, *Telomerase reverse transcriptase (TERT)* or *telomerase RNA (TERC),* are major monogenic causes of pulmonary fibrosis[65]. Patients with IPF have shortened telomeres; short telomeres[65, 66] and TERT/TERC-disease associated variants were associated with specific clinical and biological features and reduced transplant-free survival. Similarly, short telomeres increase the risk of severe COVID-19[67-69], pulmonary fibrosis and poorer outcomes[69, 70]. In the case of IL15, the presence of this cytokine in the circulation has been proposed to serve as a biomarker for prognostication in both IPF[26, 71, 72] and COVID-19[18].

Hamsters, but not mice recapitulate the host immune response and AT2 cytopathic features of COVID-19

Evidence suggests that there is a wide gap between COVID-19 in humans and animal models[73]; interspecies-related differences, such as host specificity and divergent immune responses can lead to misinterpretation. To objectively assess which model improves the translational potential of any discovery from this study, we curated all publicly available transcriptomic datasets from pre-clinical rodent models of COVID-19 and analyzed them for the induction of the ViP/sViP, IL15/IL15RA composite score, CoV-lung, IPF signatures and signatures of AT2 senescence and transition-state arrest (DATP) (**Supplementary Figure S1A**). We saw that the golden Syrian hamsters, but not the humanized transgenic mouse models (**Supplementary Figure S1A**; *bottom*) induce most, if not all the key signatures upon SARS-CoV-2 challenge (**Supplementary Figure S1A**; *top*). Most importantly, all signatures (**Supplementary Figure S1A**; *top*), including that of AT2-damage associated transient progenitor population (**Supplementary Figure S1B**; DATP) were effectively reversed with two therapeutic approaches. The first approach was the use of N-hydroxycytidine, the parent of the prodrug MK-4482 (Molnupiravir, EIDD-2801) which has not only proven as a potent and selective oral antiviral nucleoside analogue in mice, guinea pigs, ferrets and human airway epithelium organoids[74-79], but also showed promise in Phase IIa trials in the treatment of COVID-19 patients (NCT04405570[80]). both biologic and directly acting anti-viral therapies. The second approach was the use of SARS-CoV-2-neutralizing antibodies whose design was inspired by monoclonal antibodies (mAbs) isolated from from convalescent donors[28]. A specific isotype of this antibody, which binds to the receptor-binding domain (RBD-A) of SARS-CoV-2 spike protein in a fashion that precludes binding to host ACE2, was demonstrated as effective in preventing infection and weight loss symptoms, in cell-based and *in vivo* hamster models of infection, respectively.

These findings show that hamster lungs best recapitulate the host immune response, the nature of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cytokine storm (IL15-centric), and the AT2 cytopathies that are shared between COVID-19 and IPF. It is also in keeping with emerging data that Syrian hamsters best emulate COVID-19[81]; they also best emulate progressive fibrosis in IPF[82].

Validation of the predicted AT2 cytopathic features of COVID-19 in hamster and human lungs

We next sought to determine if the SARS-CoV-2 virus can induce the AT2 cytopathic changes, and whether effective therapeutics can abrogate those changes. We analyzed by RNA seq and trichrome stain the lungs from SARS-CoV-2-challenged golden Syrian hamsters who were pre-treated either with EIDD-2801 or anti-Spike (CoV-2) mAb or untreated controls (see study protocol in **Figure 5A**). Key genes associated with AT2 senescence (*Tp53* and *Cdkn2a*), ER stress (*Hspa5/Grp78*), DATP (*Krt8* and *Cldn4*) and fibrosis (*Col1a, Col3a* and *Serpine1*) were all induced in the infected lungs (compared to uninfected controls) but not in those that were treated (**Figure 5B**). Trichrome staining confirmed that compared to uninfected controls, collagen deposition was induced in the infected lungs, but not in those that were treated (**Figure 5C-D**). Immunohistology studies confirmed that Cdkn2a, Krt8, Cldn4 and Tp53 proteins were expressed in the injured alveoli, but not in the treated lungs (**Figure 5E-F**).

We next confirmed extensive deposition of eosin-positive collagenous materials in lung tissues obtained at rapid autopsies on deceased subjects with COVID-19 compared to normal lung tissues (**Figure 6A**; *top*). KRT8 and TP53 were also induced in the COVID-19 lungs (**Figure 6A**; *middle* and *bottom*). Finally, plasma ELISA studies in a cohort of symptomatic COVID-19 patients who presented to the UC San Diego Medical Center with varying disease severity, ranging from mild to fatal (see **Supplemental Information 3-4**), we detected several cellular senescence-associated cytokines during the acute compared to the convalescent visit (**Figure 6B; Supplementary Figure S2**). It is noteworthy that VEGF, TNFα, IL6 and IL1β, all cytokines that have been implicated in the fibrotic process post-COVID-19 (reviewed in[83]) were significantly elevated only in the most critical form of the disease. Findings are also consistent with the fact that VEGF-targeted therapeutics (Bevacizumab) in COVD-19 patients improved lung oxygenation up to day 28 during follow-up[84].

We next asked if the key aspects of the host immune response and AT2 cytopathic changes are induced in the various human pre-clinical *in vitro* models of COVID-19. We assessed our published models of COVID-19-in-a-dish[27] (**Figure 6C, 6D-***top 4 rows*) for AT2-senescence and for their ability to mimic the unique milieu in COVID-19 lungs[13]. We also analyzed CoV2-challenged primary(p)AT2 and A549 cells (**Figure 6D-***bottom 3 rows*). We found that submerged monolayers of adult lung organoids (ALO mono), which were previously found to outperform fetal lung organoids[85] and other models in recapitulating the gene expression signatures observed in human COVID-19 lungs[27], also outperformed here. While most models induced some signatures, the submerged ALO model induced them all (**Figure 6D**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Protein-protein network analysis predicts AT2 ER stress as an early initiating event in COVID-19

To identify some of the earliest AT2 cell processes that might be triggered as a consequence of the host immune response in acute COVID-19, which may precede the development of transient state arrest and senescence, we resorted to a PPI network analysis. To this end, we used the genes in the sViP[18], AT2 senescence[63] and DATP[60] and DATP[60] signatures as 'seed' nodes to fetch additional connecting nodes from human protein-protein interactions in STRING database[19] (**Figure 7A;** see **Supplemental Information 5)**. Once built, we assessed the network for degrees of connectivity (**Figure 7B**; *left*) and representation of cellular processes in the most connected nodes by using two approaches: CluGo analysis[86] (**Figure 7B**; *right*) and Reactome pathway analysis[87] (**Figure 7C**). While CluGO analysis showed an overrepresentation of proteins associated unfolded protein response, ER stress, regulation of telomerase, regulation of cell aging and regulation of G1/S phase transition (**Figure 7B**; *right*), reactome analysis showed enrichment of pathways associated with senescence (**Figure 7C**; *red*), platelet activation and degranulation (**Figure 7C**; *blue*), and Tp53-dependent DNA damage response (**Figure 7B**; *green*). That platelet activation (indicative of a prothrombotic state) is captured within the PPI network is in keeping with what appears to be a hallmark of PCLD[13]. Numerous independent observations in diverse cohorts of patients have now confirmed that platelet dysfunction amplifies endotheliopathy in COVID-19[88] and that a vicious cycle of thromboinflammation, endothelial dysfunction and immunothrombosis are key pathogenic mechanisms in COVID-19[89, 90]. It is noteworthy, that we found that levels of IL6 and platelets showed significant negative correlation (**Supplementary Figure 2**). Because IL6 has been specifically implicated in platelet hyper-reactivity and thrombosis[91], as well as in thrombus resolution[92], the strong negative correlation may suggest the presence of thromboinflammation; the latter is one of the first and most distinctive pathological feature of CoV-lung to be reported since the beginning of this pandemic[13]. As for AT2-centric processes, the PPI network analyses suggest that ER stress, which has been shown to serve as a causal role for profibrogenic epithelial states (SASP) that is shared between aging-associated pulmonary fibrosis and IPF[93], may also be a common underlying factor in fibrotic PCLD.

ER stress in AT2 cells is observed in COVID-19 and is sufficient to recapitulate the host immune response in COVID-19 and IPF. We next sought to investigate if alveolar ER stress is a feature in CoV-lung. To this end, we analyzed infected human (**Figure 8A**) and hamster (**Figure 8B**) lungs for the ER chaperone *Grp78*, a key regulator of ER homeostasis expression by IHC. We found that compared to normal uninfected lungs, infected lungs showed a significant increase in Grp78 staining and that the intensity of such staining was highest in the epithelial cells lining the alveolar spaces in both species (**Figure 8A-B**).

To determine if ER stress in AT2 cells may contribute to the epithelial dysfunction and fibrosis that is observed as a sequel of COVID-19, we leveraged a previously characterized *tg*-mouse model of AT2-specific

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conditional [tamoxifen (Tmx)-inducible] deletion of Grp78 (*Sftpc*$^{+/creERT2}$;*Grp78*$^{f/f}$)[93]. These mice are known to lose weight and die (more pronounced in male and old mice) due to lung inflammation, and spatially heterogeneous fibrosis characterized by fibroblastic foci and hyperplastic AT2 cells, all features of IPF[93]. We found that compared to their littermate controls, lungs from these mice (unlike other transgenic mouse models of COVID-19; see **Figure S1A**) induced the ViP/sViP, CoV2-lung and IPF signatures, as well as the AT2 cytopathic changes such as DATP and AT2 senescence (**Figure 8C**; *bottom row*). In fact, the pattern of gene expression was similar to SARS-CoV-2-challenged hamster lungs (**Figure 8C**; *top row*). These findings indicate that ER stress is sufficient to faithfully recapitulate the host immune response and alveolar cytopathic changes that are observed in PCLD and IPF.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DISCUSSION

Using an integrated computational and experimental approach (see study design, **Figure 1**), the major discovery we report here is that PCLD resembles IPF, the most common form of ILD, at a fundamental level—showing similar gene expression patterns (ViP and IPF signatures), dysfunctional cell states (AT2 and monocytes), and dysfunctional AT2 processes (ER stress, telomere instability and senescence)]. Our findings support the following model (summarized as schematic in **Figure 8D**): Upon injury (viral infection/inflammatory cytokines), diffuse alveolar damage is associated with extensive loss of AT1 cells, which are duly replaced by AT2 cells serve as progenitors that are capable of self-renewal and differentiation into AT1 cells[94-96]. But such physiologic regeneration and healing process is impaired when AT2 progenitors are also injured and begin to elicit ER stress responses, which induces telomerase activity[97, 98]. Because severe COVID-19 has been linked to short telomere lengths[98] [not as a cause, i.e, SARS-CoV-2 infection does not shorten telomeres, but as a predisposition[67]], it is possible that patients who have short and/or dysfunctional telomeres fail to adequately respond to ER stress, and instead, become senescent (through p53 activation), accumulate DNA damage and enter a dysfunctional *KRT8*+ transitional stem cell state[99]. The latter is a phenomenon that has been independently reported by three groups, under different names: PATS, pre-alveolar type-1 transitional cell state[62]; ADI, alveolar differentiation intermediates[61]; DATP, damage-associated transient progenitors[60]. All three groups reported their presence in fibrotic regions of IPF lungs; an increase of AT2 cells in the transitional state was invariably accompanied by an increase in myofibroblasts. Our PPI network analyses suggested that AT2 senescence and progenitor arrested state in the setting of a ViP/sViP-immune response (cytokine storm contributed by PBMCs) may support a SASP phenotype, which is a pathological feature of aging and IPF lung[100, 101]. It is possible that fibroblasts/myofibroblasts also contribute to this vicious cycle of inflammation and AT2 dysfunction. These findings show that these two distinct clinical syndromes, IPF, which predates the current pandemic by many decades, and the novel COVID-19, share a similar profile of host immune response, both in the lung microenvironment (in AT2 cells, to be specific) as well as in the circulating blood/PBMCs. We not only formally define the nature of that host response and provide computational tools to measure the extent of such response, but also chart the cascade of cytopathic changes in the alveoli that are critical for the profibrogenic state. It is noteworthy that the same host immune response is seen also in MIS-C, a new disease that co-emerged with COVID-19, and in KD (which shares overlapping features with MIS-C in clinical presentation[102]).

There are three major impacts and/implications of the findings reported in this study. First, our finding that COVID-19 and IPF share fundamental host immune response and alveolar cytopathic features is in keeping with the fact that both diseases share epidemiologic similarities—both primarily occurs in older adults and is characterized by progressive worsening of dyspnea and lung function[103, 104]. Prevalence of both increases with

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

age, with most patients aged >50 years at diagnosis and a higher proportion of males than females. Risk factors reported to be associated with IPF include cigarette smoking, environmental exposures, microbial pathogens, and genetic risk factors[103]. Findings also suggest that IPF may be a multi-trigger infection-driven chronic inflammatory condition, as has been suggested in the past decade[32, 105-107]. Finally, our findings are in keeping with epidemiologic evidence for an increased risk of death in patients with existing ILDs from COVID-19 compared with adults without ILD, even after controlling for age, sex, and comorbidities[108-110]. At day 30 of COVID-19, 35% of patients with fibrotic idiopathic ILD had died compared with 19% of those with other ILDs[111]. Whether the shared molecular or pathophysiologic features we report here synergize to increase fatality remains unknown.

Second, most severe COVID-19 patients develop pneumonia and hyperinflammation likely related to a macrophage activation syndrome[112] commonly named "cytokine storm". Linked to the storm, lung fibrosis emerges as a secondary event related to the progression of the pathology[1]. Although this storm has been implicated[113], and it is comprised of the generic mix of all the typical cytokines[10], which exact component is the pro-fibrogenic driver remains unclear. By showing that a IL15/IL15RA-centric cytokine storm is the key shared phenomenon, this study provides a link between the post-viral hyperinflammation phase and the sequelae of fibrosis. The presence of IL15 in the circulation has been proposed to serve as a biomarker for prognostication in IPF and other ILDs[114-116]. As for what may be contributing to such a cytokine response, myeloid cells are likely to be major culprits, but AT2 cells cannot be ruled out. Prior studies using bleomycin-challenged transgenic mice that lacks the *Telomeric Repeat Binding Factor 2* in the AT2 cells (*Trf2*$^{Fl/Fl}$*;Sftpc-CreER*) have shown that AT2 cells with telomere dysfunction upregulate an IL15-centric pathways[117]. It is possible that injury, DNA damage and IL15 signalling are one component in the profibrogenic cascade and suggest that IL15-targeted therapeutics may be beneficial in the most severe cases of COVID-19 to prevent fibrotic sequelae. Finally, our findings support the rationalization of clinical trials in COVID-19 using FDA-approved drugs for IPF, nintedanib (Ofev®) and pirfenidone (Esbriet®) (NCT04856111 and NCT04653831), and anecdotal case reports and case series have already chronicled the benefits of their use in COVID-19 lung disease[118-120].

Third, our finding that shared alveolar cytopathic changes (e.g., DNA damage, progenitor state arrest, SASP, and ER stress) fuel fibrogenic programs in both COVID-19 and IPF are insightful because AT2 cells are known to contain an elegant quality control feedback loop to respond to intrinsic or extrinsic stress; a failure of such quality control results in diverse cellular phenotypes (reviewed in [121]): ER stress [122], defective autophagy, mitochondrial dysfunction, apoptosis, inflammatory cell recruitment, profibrotic signaling, and altered progenitor function[99, 123, 124] that ultimately converge to drive downstream fibrotic remodeling in the lung. Prior work[63],

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

which led to the discovery of the AT2-senescence signature (which we used here), had demonstrated that senescence of AT2 cells (not loss of AT2) is sufficient to drive progressive pulmonary fibrosis. Others agree, and this is now an established pathophysiologic trigger in lung fibrosis[57, 125-127]. Our findings are in keeping with published work[128] that suggest that AT2 senescence may be a targetable disease driver of lung injury in COVID-19. Although AT2 senescence is a shared phenomenon, our PPI network analyses --which integrated AT2 processes with the immune responses (ViP signatures) -- provided valuable clues into how platelet activation and thromboinflammation may be uniquely seen in the setting of COVID-19. These findings are in keeping with recent publications[129-131] which suggest that SARS-CoV-2 could induce epithelial senescence; senescencet AT2 cells would then assume a SASP phenotype, which in turn led to neutrophil and platelet activation, and activation of the clotting cascade[129].

As for the limitations of this study, it remains unclear how to model progressive lung fibrosis *in vitro* and hence, no attempt was made to do so. Our results suggest that AT2 cells alone may be insufficient for such modeling because a specific host immune response that is carried in the PBMCs is a clear determinant as to who progresses and who does not. Although AT2-specific modulation of ER-stress pathway and SARS-CoV-challenged (treated vs untreated) hamsters were used to go beyond association and establish causation, our study did not attempt to inhibit/reverse fibrosis in COVID-19 by acting on any profibrogenic cellular pathway/process. Development of novel chemical matter/biologicals and validation of the therapeutic efficacy of such agents will take time, but if successful, our findings show that their benefits will likely transcend beyond PCLD into IPF and other fibrotic lung conditions such as IPF.

In conclusion, this transdisciplinary work provides insights into the pathogenesis of PCLD, formally defines the fibrogenic processes in the lung, and rigorously validates high value gene signatures or even targets (i.e., IL15, senescence pathways, etc.) to track and manipulate the same. The insights, tools, computationally vetted disease models and biomarkers (prognostic gene signatures) identified here are likely to spur the development of therapies for patients with fibrotic interstitial lung disease of diverse causes, including IPF, all of whom have limited or no good treatment options.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

AUTHOR CONTRIBUTIONS

DS, SS and PG conceptualized the project; SS and DS carried out all computational modeling and analyses and DS contributed all software used in this work; GDK carried out the human serum cytokine analysis; VC and SS carried out the immunohistochemical studies and their quantitative analysis under the supervision of PG; AGF carried out the immunofluorescence studies on COVID-19 lung tissues; SS carried out the PPI network creation and analysis; JD provided access to human subjects for cytokine analyses. CT, CRE and SD were responsible for the COVID-19 lung models. SS, DS and PG prepared figures for data visualization; SS and PG wrote the original draft of the manuscript; DS, SD, GK edited and revised the manuscript. All co-authors approved the final version of the manuscript. DS, SD, and PG supervised various parts of the project and secured funding; DS and PG administered the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes for Health (NIH) grants R00-CA151673 and R01-GM138385 (to DS) and R01-AI141630, CA100768 and CA160911 (to P.G), R01DK107585 (SD) and R01-AI155696 (to P.G, D.S and S.D), UCOP-RGPO (R00RG2628 & R00RG2642 to P.G, D.S and S.D). GDK was supported through the American Association of Immunologists Intersect Fellowship Program for Computational Scientists and Immunologists. We thank Zea Borok (UCSD) for helpful discussions during the preparation of this manuscript.

Data sharing:

All data is available in the main text or the supplementary materials.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUPPLEMENTARY MATERIALS:

- Detailed methods (text)
- Figures **2**
- Tables **0**
- Supplemental datasheets– **1-5**
- **Supplemental Information 1**: Excel datasheet with a list of transcriptomic datasets that were used in this work.
- **Supplemental Information 2**: Excel datasheet with gene expression signatures that were used in this work.
- **Supplemental Information 3**: COVID-19 patient demographics whose plasma were analyzed for cytokines.
- **Supplemental Information 4**: Cytokine expression profiles in COVID-19 patients vs healthy controls.
- **Supplemental Information 5**: Excel datasheet with list of degree sorted nodes in the protein-protein interaction network in **Fig 7B**.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Spagnolo, P. et al. Pulmonary fibrosis secondary to COVID-19: a call to arms? *Lancet Respir Med* **8**, 750-752 (2020).
2. Myall, K.J. et al. Persistent Post-COVID-19 Interstitial Lung Disease. An Observational Study of Corticosteroid Treatment. *Ann Am Thorac Soc* **18**, 799-806 (2021).
3. Wells, A.U., Devaraj, A. & Desai, S.R. Interstitial Lung Disease after COVID-19 Infection: A Catalog of Uncertainties. *Radiology* **299**, E216-e218 (2021).
4. Aronson, K.I. & Podolanczuk, A.J. Lungs after COVID-19: Evolving Knowledge of Post-COVID-19 Interstitial Lung Disease. *Ann Am Thorac Soc* **18**, 773-774 (2021).
5. Dadhwal, R., Sharma, M. & Surani, S. Restrictive Lung Disease in Patients With Subclinical Coronavirus Infection: Are We Bracing Ourselves for Devastating Sequelae? *Cureus* **13**, e12501 (2021).
6. Atabati, E., Dehghani-Samani, A. & Mortazavimoghaddam, S.G. Association of COVID-19 and other viral infections with interstitial lung diseases, pulmonary fibrosis, and pulmonary hypertension: A narrative review. *Can J Respir Ther* **56**, 1-9 (2020).
7. Ahmad Alhiyari, M. et al. Post COVID-19 fibrosis, an emerging complicationof SARS-CoV-2 infection. *IDCases* **23**, e01041 (2021).
8. Alharthy, A. et al. Residual Lung Injury in Patients Recovering From COVID-19 Critical Illness: A Prospective Longitudinal Point-of-Care Lung Ultrasound Study. *J Ultrasound Med* (2020).
9. Gagiannis, D. et al. Clinical, Serological, and Histopathological Similarities Between Severe COVID-19 and Acute Exacerbation of Connective Tissue Disease-Associated Interstitial Lung Disease (CTD-ILD). *Front Immunol* **11**, 587517 (2020).
10. George, P.M., Wells, A.U. & Jenkins, R.G. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *Lancet Respir Med* **8**, 807-815 (2020).
11. van Gassel, R.J.J. et al. Functional Outcomes and Their Association With Physical Performance in Mechanically Ventilated Coronavirus Disease 2019 Survivors at 3 Months Following Hospital Discharge: A Cohort Study. *Crit Care Med* (2021).
12. Kaarteenaho, R. & Kinnula, V.L. Diffuse alveolar damage: a common phenomenon in progressive interstitial lung disorders. *Pulm Med* **2011**, 531302 (2011).
13. Ackermann, M. et al. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *N Engl J Med* **383**, 120-128 (2020).
14. Cheon, I.S. et al. Immune signatures underlying post-acute COVID-19 lung sequelae. *Sci Immunol*, eabk1741 (2021).
15. Bharat, A. et al. Lung transplantation for patients with severe COVID-19. *Sci Transl Med* **12** (2020).
16. Croci, G.A. et al. Emergency Lung Transplantation after COVID-19: Immunopathological Insights on Two Affected Patients. *Cells* **10** (2021).
17. Bharat, A. et al. Early outcomes after lung transplantation for severe COVID-19: a series of the first consecutive cases from four countries. *Lancet Respir Med* **9**, 487-497 (2021).
18. Sahoo, D. et al. AI-guided discovery of the invariant host response to viral pandemics. *EBioMedicine* **68**, 103390 (2021).
19. Szklarczyk, D. et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* **49**, D605-D612 (2021).
20. Fink, G. et al. Inactivated trivalent influenza vaccination is associated with lower mortality among patients with COVID-19 in Brazil. *BMJ Evid Based Med* (2020).
21. Wagstaffe, H.R. et al. Influenza Vaccination Primes Human Myeloid Cell Cytokine Secretion and NK Cell Function. *J Immunol* **203**, 1609-1618 (2019).
22. Hoek, K.L. et al. A cell-based systems biology assessment of human blood to monitor immune responses after influenza vaccination. *PLoS One* **10**, e0118528 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

23. Carson, W.E. et al. Interleukin (IL) 15 is a novel cytokine that activates human natural killer cells via components of the IL-2 receptor. *J Exp Med* **180**, 1395-1403 (1994).
24. Carson, W.E. et al. Endogenous production of interleukin 15 by activated human monocytes is critical for optimal production of interferon-gamma by natural killer cells in vitro. *J Clin Invest* **96**, 2578-2582 (1995).
25. Sahoo, D., Dill, D.L., Gentles, A.J., Tibshirani, R. & Plevritis, S.K. Boolean implication networks derived from large scale, whole genome microarray datasets. *Genome Biol* **9**, R157 (2008).
26. Sahoo, D. et al. MiDReG: a method of mining developmentally regulated genes using Boolean implications. *Proc Natl Acad Sci U S A* **107**, 5732-5737 (2010).
27. Tindle, C. et al. Adult stem cell-derived complete lung organoid models emulate lung disease in COVID-19. *Elife* **10** (2021).
28. Rogers, T.F. et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* **369**, 956-963 (2020).
29. Varghese, F., Bukhari, A.B., Malhotra, R. & De, A. IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples. *PLoS One* **9**, e96801 (2014).
30. Levin, M. Childhood Multisystem Inflammatory Syndrome - A New Challenge in the Pandemic. *N Engl J Med* **383**, 393-395 (2020).
31. Whittaker, E. et al. Clinical Characteristics of 58 Children With a Pediatric Inflammatory Multisystem Syndrome Temporally Associated With SARS-CoV-2. *Jama* **324**, 259-269 (2020).
32. Imai, Y. et al. Angiotensin-converting enzyme 2 protects from severe acute lung failure. *Nature* **436**, 112-116 (2005).
33. Bonham, C.A., Strek, M.E. & Patterson, K.C. From granuloma to fibrosis: sarcoidosis associated pulmonary fibrosis. *Curr Opin Pulm Med* **22**, 484-491 (2016).
34. Patterson, K.C. & Strek, M.E. Pulmonary fibrosis in sarcoidosis. Clinical features and outcomes. *Ann Am Thorac Soc* **10**, 362-370 (2013).
35. Meltzer, E.B. et al. Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. *BMC Med Genomics* **4**, 70 (2011).
36. Bauer, Y. et al. A novel genomic signature with translational significance for human idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* **52**, 217-231 (2015).
37. Gangwar, I. et al. Detecting the Molecular System Signatures of Idiopathic Pulmonary Fibrosis through Integrated Genomic Analysis. *Sci Rep* **7**, 1554 (2017).
38. Ammar, R., Sivakumar, P., Jarai, G. & Thompson, J.R. A robust data-driven genomic signature for idiopathic pulmonary fibrosis with applications for translational model selection. *PLoS One* **14**, e0215565 (2019).
39. Wu, X. et al. 3-month, 6-month, 9-month, and 12-month respiratory outcomes in patients following COVID-19-related hospitalisation: a prospective study. *Lancet Respir Med* **9**, 747-754 (2021).
40. Marshall, M. The lasting misery of coronavirus long-haulers. *Nature* **585**, 339-341 (2020).
41. Kølner-Augustson, L., Prior, T.S., Skivild, V., Aalestrup, A. & Bendstrup, E. Fatigue in idiopathic pulmonary fibrosis measured by the Fatigue Assessment Scale during antifibrotic treatment. *Eur Clin Respir J* **8**, 1853658 (2020).
42. Bloem, A.E.M. et al. Severe Fatigue is Highly Prevalent in Patients with IPF or Sarcoidosis. *J Clin Med* **9** (2020).
43. Bloem, A.E.M. et al. Perceptions of fatigue in patients with idiopathic pulmonary fibrosis or sarcoidosis. *J Thorac Dis* **13**, 4872-4884 (2021).
44. Vasa, C.V. State of the Globe: The H1N1 Threat Continues to Loom the Planet. *J Glob Infect Dis* **5**, 91-92 (2013).
45. van der Meer, F.J., Orsel, K. & Barkema, H.W. The new influenza A H1N1 virus: balancing on the interface of humans and animals. *Can Vet J* **51**, 56-62 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

46. Bermejo-Martin, J.F. et al. Host adaptive immunity deficiency in severe pandemic influenza. *Crit Care* **14**, R167 (2010).
47. Baden, L.R. et al. H1N1 influenza A disease--information for health professionals. *N Engl J Med* **360**, 2666-2667 (2009).
48. Zhong, N.S. et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353-1358 (2003).
49. Beigel, J.H. et al. Avian influenza A (H5N1) infection in humans. *N Engl J Med* **353**, 1374-1385 (2005).
50. Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D. & Fouchier, R.A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* **367**, 1814-1820 (2012).
51. Zhai, Y. et al. Host Transcriptional Response to Influenza and Other Acute Respiratory Viral Infections--A Prospective Cohort Study. *PLoS Pathog* **11**, e1004869 (2015).
52. Leahy, T.R. et al. Interleukin-15 is associated with disease severity in viral bronchiolitis. *Eur Respir J* **47**, 212-222 (2016).
53. Evren, E. et al. Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity* **54**, 259-275.e257 (2021).
54. Zhang, Z. et al. SCINA: A Semi-Supervised Subtyping Algorithm of Single Cells and Bulk Samples. *Genes (Basel)* **10** (2019).
55. Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* **26**, 842-844 (2020).
56. Wu, K. et al. TREM-2 promotes macrophage survival and lung disease after respiratory viral infection. *J Exp Med* **212**, 681-697 (2015).
57. Parimon, T., Yao, C., Stripp, B.R., Noble, P.W. & Chen, P. Alveolar Epithelial Type II Cells as Drivers of Lung Fibrosis in Idiopathic Pulmonary Fibrosis. *Int J Mol Sci* **21** (2020).
58. Garcia, C.K. Insights from human genetic studies of lung and organ fibrosis. *J Clin Invest* **128**, 36-44 (2018).
59. Pardo, A. & Selman, M. Molecular mechanisms of pulmonary fibrosis. *Front Biosci* **7**, d1743-1761 (2002).
60. Choi, J. et al. Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration. *Cell Stem Cell* **27**, 366-382.e367 (2020).
61. Strunz, M. et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat Commun* **11**, 3559 (2020).
62. Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat Cell Biol* **22**, 934-946 (2020).
63. Yao, C. et al. Senescence of Alveolar Type 2 Cells Drives Progressive Pulmonary Fibrosis. *Am J Respir Crit Care Med* **203**, 707-717 (2021).
64. Sahoo, D., Dill, D.L., Tibshirani, R. & Plevritis, S.K. Extracting binary signals from microarray time-course data. *Nucleic Acids Res* **35**, 3705-3712 (2007).
65. Abdul-Careem, M.F. et al. Critical role of natural killer cells in lung immunopathology during influenza infection in mice. *J Infect Dis* **206**, 167-177 (2012).
66. Schneider, W.M., Chevillotte, M.D. & Rice, C.M. Interferon-stimulated genes: a complex web of host defenses. *Annu Rev Immunol* **32**, 513-545 (2014).
67. Aviv, A. Short telomeres and severe COVID-19: The connection conundrum. *EBioMedicine* **70**, 103513 (2021).
68. Froidure, A. et al. Short telomeres increase the risk of severe COVID-19. *Aging (Albany NY)* **12**, 19911-19922 (2020).
69. McGroder, C.F. et al. Pulmonary fibrosis 4 months after COVID-19 is associated with severity of illness and blood leucocyte telomere length. *Thorax* (2021).
70. Mongelli, A. et al. Evidence for Biological Age Acceleration and Telomere Shortening in COVID-19 Survivors. *Int J Mol Sci* **22** (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

71. Dabydeen, S.A., Desai, A. & Sahoo, D. Unbiased Boolean analysis of public gene expression data for cell cycle gene identification. *Mol Biol Cell* **30**, 1770-1779 (2019).
72. Sodhi, C.P. et al. Attenuation of pulmonary ACE2 activity impairs inactivation of des-Arg. *Am J Physiol Lung Cell Mol Physiol* **314**, L17-L31 (2018).
73. Abdelnabi, R. et al. The combined treatment of Molnupiravir and Favipiravir results in a potentiation of antiviral efficacy in a SARS-CoV-2 hamster infection model. *EBioMedicine* **72**, 103595 (2021).
74. Cox, R.M., Wolf, J.D. & Plemper, R.K. Therapeutically administered ribonucleoside analogue MK-4482/EIDD-2801 blocks SARS-CoV-2 transmission in ferrets. *Nat Microbiol* **6**, 11-18 (2021).
75. Toots, M. et al. Characterization of orally efficacious influenza drug with high resistance barrier in ferrets and human airway epithelia. *Sci Transl Med* **11** (2019).
76. Toots, M. et al. Quantitative efficacy paradigms of the influenza clinical drug candidate EIDD-2801 in the ferret model. *Transl Res* **218**, 16-28 (2020).
77. Sheahan, T.P. et al. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci Transl Med* **12** (2020).
78. Urakova, N. et al. β-d-N (4)-Hydroxycytidine Is a Potent Anti-alphavirus Compound That Induces a High Level of Mutations in the Viral Genome. *J Virol* **92** (2018).
79. Abdelnabi, R. et al. Molnupiravir (EIDD-2801) inhibits SARS-CoV-2 replication and enhances the efficacy of favipiravir in a Syrian hamster infection model. *bioRxiv*, 2020.2012.2010.419242 (2021).
80. Fischer, W. et al. Molnupiravir, an Oral Antiviral Treatment for COVID-19. *medRxiv* (2021).
81. Chang, E.H. et al. Rhinovirus Infections in Individuals with Asthma Increase ACE2 Expression and Cytokine Pathways Implicated in COVID-19. *Am J Respir Crit Care Med* **202**, 753-755 (2020).
82. Yang, P. et al. Angiotensin-converting enzyme 2 (ACE2) mediates influenza H7N9 virus-induced acute lung injury. *Sci Rep* **4**, 7027 (2014).
83. Vasarmidi, E., Tsitoura, E., Spandidos, D.A., Tzanakis, N. & Antoniou, K.M. Pulmonary fibrosis in the aftermath of the COVID-19 era (Review). *Exp Ther Med* **20**, 2557-2560 (2020).
84. Pang, J. et al. Efficacy and tolerability of bevacizumab in patients with severe Covid-19. *Nat Commun* **12**, 814 (2021).
85. Lamers, M.M. et al. An organoid-derived bronchioalveolar model for SARS-CoV-2 infection of human alveolar type II-like cells. *EMBO J* **40**, e105912 (2021).
86. In-Yee Lee and Jan-Ming Ho and Ming-Syan, C. 4 - (2005).
87. Fabregat, A. et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **46**, D649-D655 (2018).
88. Barrett, T.J. et al. Platelets amplify endotheliopathy in COVID-19. *Sci Adv* **7**, eabh2434 (2021).
89. Bonaventura, A. et al. Endothelial dysfunction and immunothrombosis as key pathogenic mechanisms in COVID-19. *Nat Rev Immunol* **21**, 319-329 (2021).
90. Maruhashi, T. & Higashi, Y. Pathophysiological Association of Endothelial Dysfunction with Fatal Outcome in COVID-19. *Int J Mol Sci* **22** (2021).
91. Senchenkova, E.Y. et al. Interleukin-6 mediates the platelet abnormalities and thrombogenesis associated with experimental colitis. *Am J Pathol* **183**, 173-181 (2013).
92. Nosaka, M. et al. Crucial Involvement of IL-6 in Thrombus Resolution in Mice via Macrophage Recruitment and the Induction of Proteolytic Enzymes. *Front Immunol* **10**, 3150 (2019).
93. Borok, Z. et al. Loss in Epithelial Progenitors Reveals an Age-linked Role for Endoplasmic Reticulum Stress in Pulmonary Fibrosis. *Am J Respir Crit Care Med* **201**, 198-211 (2020).
94. Rock, J.R. et al. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc Natl Acad Sci U S A* **108**, E1475-1483 (2011).
95. Barkauskas, C.E. et al. Type 2 alveolar cells are stem cells in adult lung. *J Clin Invest* **123**, 3025-3036 (2013).
96. Frank, D.B. et al. Emergence of a Wave of Wnt Signaling that Regulates Lung Alveologenesis by Controlling Epithelial Self-Renewal and Differentiation. *Cell Rep* **17**, 2312-2325 (2016).
97. Zhou, J. et al. Endoplasmic reticulum stress activates telomerase. *Aging Cell* **13**, 197-200 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

98. Sanchez-Vazquez, R., Guío-Carrión, A., Zapatero-Gaviria, A., Martínez, P. & Blasco, M.A. Shorter telomere lengths in patients with severe COVID-19 disease. *Aging (Albany NY)* **13**, 1-15 (2021).
99. Verheyden, J.M. & Sun, X. A transitional stem cell state in the lung. *Nat Cell Biol* **22**, 1025-1026 (2020).
100. Liu, R.M. & Liu, G. Cell senescence and fibrotic lung diseases. *Exp Gerontol* **132**, 110836 (2020).
101. Parimon, T., Hohmann, M.S. & Yao, C. Cellular Senescence: Pathogenic Mechanisms in Lung Fibrosis. *Int J Mol Sci* **22** (2021).
102. Sahoo, D. et al. An AI-guided signature reveals the nature of the shared proximal pathways of host immune response in MIS-C and Kawasaki disease. *bioRxiv* (2021).
103. Gaddam, R.R., Chambers, S. & Bhatia, M. ACE and ACE2 in inflammation: a tale of two enzymes. *Inflamm Allergy Drug Targets* **13**, 224-234 (2014).
104. Inlay, M.A. et al. Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev* **23**, 2376-2381 (2009).
105. Sodhi, C.P. et al. A Dynamic Variation of Pulmonary ACE2 Is Required to Modulate Neutrophilic Inflammation in Response to. *J Immunol* **203**, 3000-3012 (2019).
106. Jia, H. Pulmonary Angiotensin-Converting Enzyme 2 (ACE2) and Inflammatory Lung Disease. *Shock* **46**, 239-248 (2016).
107. Kuba, K., Imai, Y. & Penninger, J.M. Multiple functions of angiotensin-converting enzyme 2 and its relevance in cardiovascular diseases. *Circ J* **77**, 301-308 (2013).
108. Esposito, A.J. et al. Increased Odds of Death for Patients with Interstitial Lung Disease and COVID-19: A Case-Control Study. *Am J Respir Crit Care Med* **202**, 1710-1713 (2020).
109. Drake, T.M. et al. Outcome of Hospitalization for COVID-19 in Patients with Interstitial Lung Disease. An International Multicenter Study. *Am J Respir Crit Care Med* **202**, 1656-1665 (2020).
110. Podolanczuk, A.J. & Richeldi, L. COVID-19 and Interstitial Lung Disease: Keep Them Separate. *Am J Respir Crit Care Med* **202**, 1614-1616 (2020).
111. Gallay, L. et al. Risk Factors for Mortality after COVID-19 in Patients with Preexisting Interstitial Lung Disease. *Am J Respir Crit Care Med* **203**, 245-249 (2021).
112. McGonagle, D., Sharif, K., O'Regan, A. & Bridgewood, C. The Role of Cytokines including Interleukin-6 in COVID-19 induced Pneumonia and Macrophage Activation Syndrome-Like Disease. *Autoimmun Rev* **19**, 102537 (2020).
113. Garcia-Revilla, J., Deierborg, T., Venero, J.L. & Boza-Serrano, A. Hyperinflammation and Fibrosis in Severe COVID-19 Patients: Galectin-3, a Target Molecule to Consider. *Front Immunol* **11**, 2069 (2020).
114. Muro, S. et al. Expression of IL-15 in inflammatory pulmonary diseases. *J Allergy Clin Immunol* **108**, 970-975 (2001).
115. Takada, T. et al. Role of IL-15 in interstitial lung diseases in amyopathic dermatomyositis with anti-MDA-5 antibody. *Respir Med* **141**, 7-13 (2018).
116. Wuttge, D.M. et al. Serum IL-15 in patients with early systemic sclerosis: a potential novel marker of lung disease. *Arthritis Res Ther* **9**, R85 (2007).
117. Alder, J.K. et al. Telomere dysfunction causes alveolar stem cell failure. *Proc Natl Acad Sci U S A* **112**, 5099-5104 (2015).
118. FIRDAUS, H., RAZAK, N.A., REZA, M., KHAN, N.A. & SHAMEEM, M. Clinical Improvement with Pirfenidone in Two Patients of COVID 19. *Journal of Clinical & Diagnostic Research* **14** (2020).
119. Momen, A.B.I. et al. Usefulness of Pirfenidone in Covid Lung: A Case Series. *European Journal of Medical and Health Sciences* **3**, 24-26 (2021).
120. Xi, Z., Zhigang, Z. & Ting, L. Post-inflammatory pulmonary fibrosis in a discharged COVID-19 patient: Effectively treated with Pirfenidone. *Archives of Pulmonology and Respiratory Care* **6**, 051-053 (2020).
121. Katzen, J. & Beers, M.F. Contributions of alveolar epithelial cell quality control to pulmonary fibrosis. *J Clin Invest* **130**, 5088-5099 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

122. Borok, Z. et al. Grp78 Loss in Epithelial Progenitors Reveals an Age-linked Role for Endoplasmic Reticulum Stress in Pulmonary Fibrosis. *Am J Respir Crit Care Med* **201**, 198-211 (2020).
123. Liang, J. et al. Hyaluronan and TLR4 promote surfactant-protein-C-positive alveolar progenitor cell renewal and prevent severe pulmonary fibrosis in mice. *Nat Med* **22**, 1285-1293 (2016).
124. Xu, Y. et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* **1**, e90558 (2016).
125. Chilosi, M., Carloni, A., Rossi, A. & Poletti, V. Premature lung aging and cellular senescence in the pathogenesis of idiopathic pulmonary fibrosis and COPD/emphysema. *Transl Res* **162**, 156-173 (2013).
126. Chen, H. et al. TGF-β1/IL-11/MEK/ERK signaling mediates senescence-associated pulmonary fibrosis in a stress-induced premature senescence model of Bmi-1 deficiency. *Exp Mol Med* **52**, 130-151 (2020).
127. Rana, T. et al. PAI-1 Regulation of TGF-β1-induced Alveolar Type II Cell Senescence, SASP Secretion, and SASP-mediated Activation of Alveolar Macrophages. *Am J Respir Cell Mol Biol* **62**, 319-330 (2020).
128. Evangelou, K. et al. SARS-CoV-2 infects lung epithelial cells and induces senescence and an inflammatory response in patients with severe COVID-19. *bioRxiv*, 2021.2001.2002.424917 (2021).
129. Lee, S. et al. Virus-induced senescence is a driver and therapeutic target in COVID-19. *Nature* **599**, 283-289 (2021).
130. Lipskaia, L. et al. Evidence that SARS-CoV-2 Induces Lung-Cell Senescence: Potential Impact on COVID-19 Lung Disease. *Am J Respir Cell Mol Biol* (2021).
131. D'Agnillo, F. et al. Lung epithelial and endothelial damage, loss of tissue repair, inhibition of fibrinolysis, and cellular senescence in fatal COVID-19. *Sci Transl Med* **13**, eabj7790 (2021).
132. Herazo-Maya, J.D. et al. Validation of a 52-gene risk profile for outcome prediction in patients with idiopathic pulmonary fibrosis: an international, multicentre, cohort study. *Lancet Respir Med* **5**, 857-868 (2017).
133. Ju, Z. et al. Telomere dysfunction induces environmental alterations limiting hematopoietic stem cell function and engraftment. *Nat Med* **13**, 742-747 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIGURES AND LEGENDS

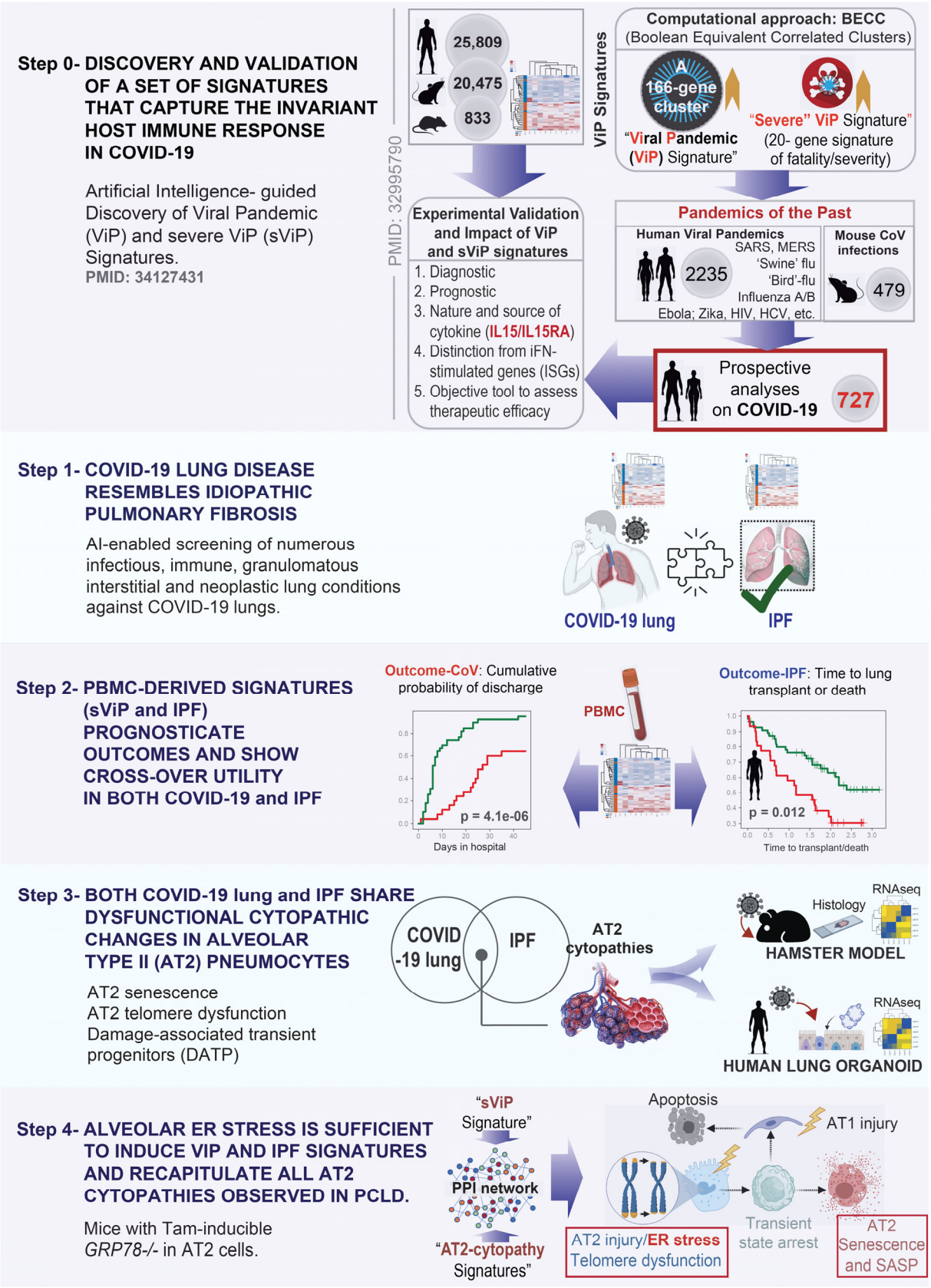

Figure 1. Study design: Artificial Intelligence-guided navigation of COVID-19 lung disease.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(*From top to bottom*) *Step 0*: Over 45,000 human, mouse, and rat gene expression databases were mined using machine learning tools called Boolean Equivalent Correlated Clusters (BECC[71]) to identify invariant host response to viral pandemics (ViP). In the absence of a sufficiently large number of COVID-19 datasets at the onset of the COVID-19 pandemic, these ViP signatures were trained on only two datasets from the pandemics of the past (Influenza and avian flu; GSE47963, n = 438; GSE113211, n = 118) and used without further training to prospectively analyze the samples from the current pandemic (i.e., COVID-19; n = 727 samples from diverse datasets). A subset of 20-genes classified disease severity called severe-ViP (sViP) signature. The ViP signatures appeared to capture the 'invariant' host response, i.e., the shared fundamental nature of the host immune response induced by all viral pandemics, including COVID-19. *Step1*: The set of ViP/sViP signatures and a CoV-lung specific[13] gene signature was analyzed on diverse transcriptomic datasets representing a plethora of lung diseases; these efforts identified COVID-19 lung disease to be the closest to Idiopathic pulmonary fibrosis (IPF); both conditions induced a common array of gene signatures. *Step 2*: Clinically useful whole-blood and PBMC-derived prognostic signatures previously validated in IPF[18] showed crossover efficacy in COVID-19, and *vice versa*. *Step 3*: Gene signatures of alveolar type II (AT2) cytopathic changes that are known to fuel IPF were analyzed in COVID-19 lung, and predicted shared features were validated in human and hamster lungs and lung-organoid derived models. *Step 4:* Protein- protein interaction (PPI) network built using sViP and AT2 cytopathy-related signatures was analyzed to pinpoint ER stress as a major shared feature in COVID-19 lung disease and IPF, which was subsequently validated in human and hamster lungs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

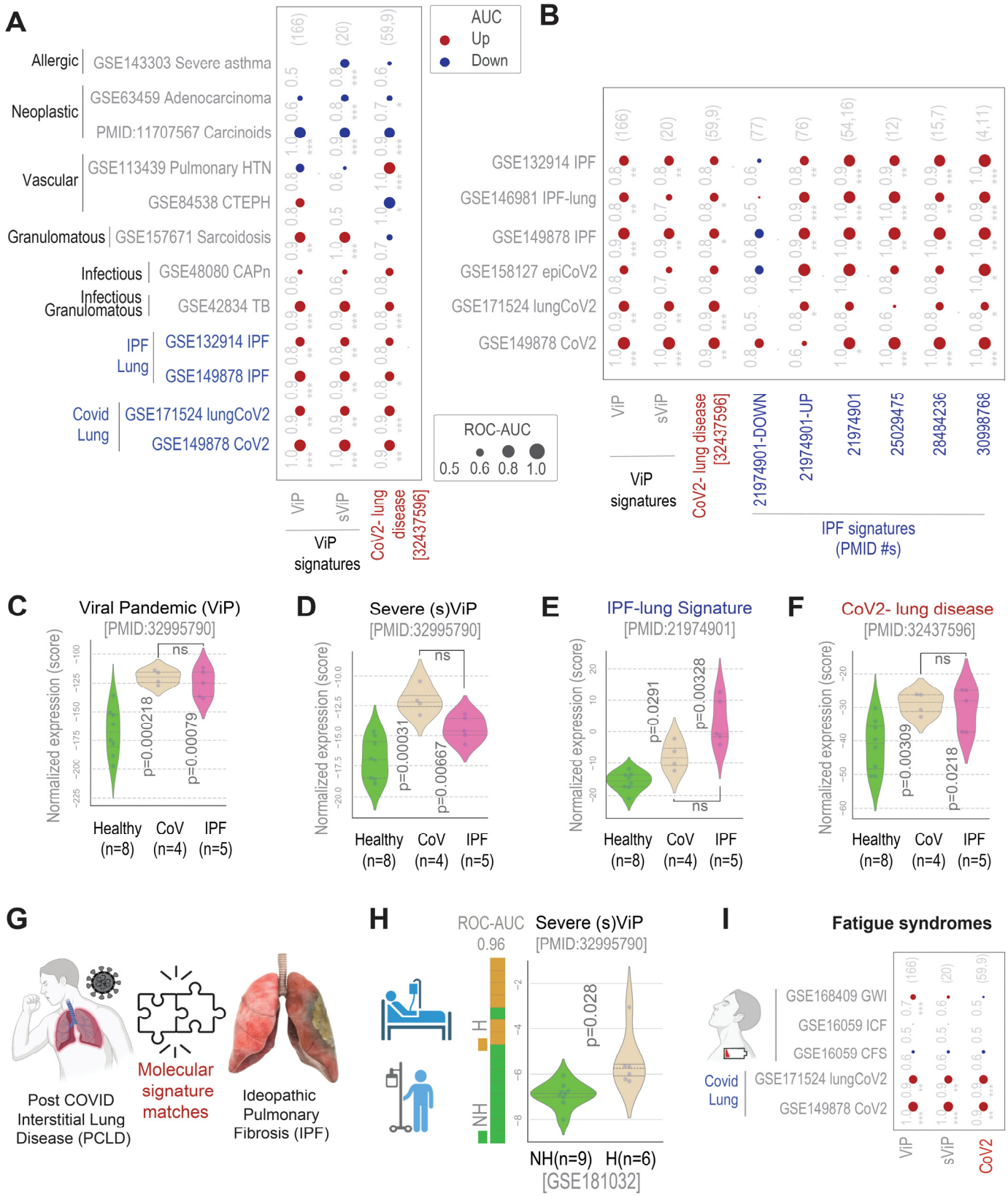

Figure 2. Identification of lung condition(s) that resemble COVID-19 lung disease.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A. Bubble plots show the ROC-AUC values (radii of circles are based on the ROC-AUC) show the direction of gene regulation (Up, red; Down, blue) for sample classification based on the induction of ViP/sViP[18] and a previously published human CoV-lung signatures[13] in diverse datasets representing diverse lung pathologies, including COVID-19. Tuberculosis, TB; community acquired pneumonia, CAPn; pulmonary hypertension, PHT; chronic thromboembolic pulmonary hypertension, CTEPH; idiopathic pulmonary fibrosis, IPF; CoV, COVID-19. **B.** Bubble plots in B show the ROC-AUC values for sample classification based on the induction of ViP signatures and several previously published IPF signatures in COVID-19 infected lungs and IPF samples. **C-F.** Violin plots display the extent of induction of ViP (E) or sViP (F), IPF (G) or CoV lung (H) signatures in a pooled cohort of healthy, CoV and IPF lung samples (GSE149878; GSE122960). Welch's two sample unpaired t-test is performed on the composite gene signature score (z-score of normalized tpm count) to compute the *p* values. In multi-group setting each group is compared to the first control group and only significant *p* values are displayed. **G.** Schematic summarizes the predicted molecular match between COVID-19 lung and IPF based on the crossover validation of gene signatures showcased in B-F. **H.** Bar plots show the ability of sViP signature to classify COVID-19 patients based on hospitalization status and violin plots show the degree of induction of the signature within each group (H, hospitalized; NH, non-hospitalized). Welch's two sample unpaired t-test is performed on the composite gene signature score (z-score of normalized tpm count) to compute the *p values*. **I.** Bubble plots of ROC-AUC values (radii of circles are based on the ROC-AUC) show the direction of gene regulation (Up, red; Down, blue) for the classification of healthy vs diseased samples based on the ViP/sViP and the CoV-lung signatures on three well known fatigue conditions (blood/PBMC samples from GWI, gulf war illness; CFS, chronic fatigue syndrome).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

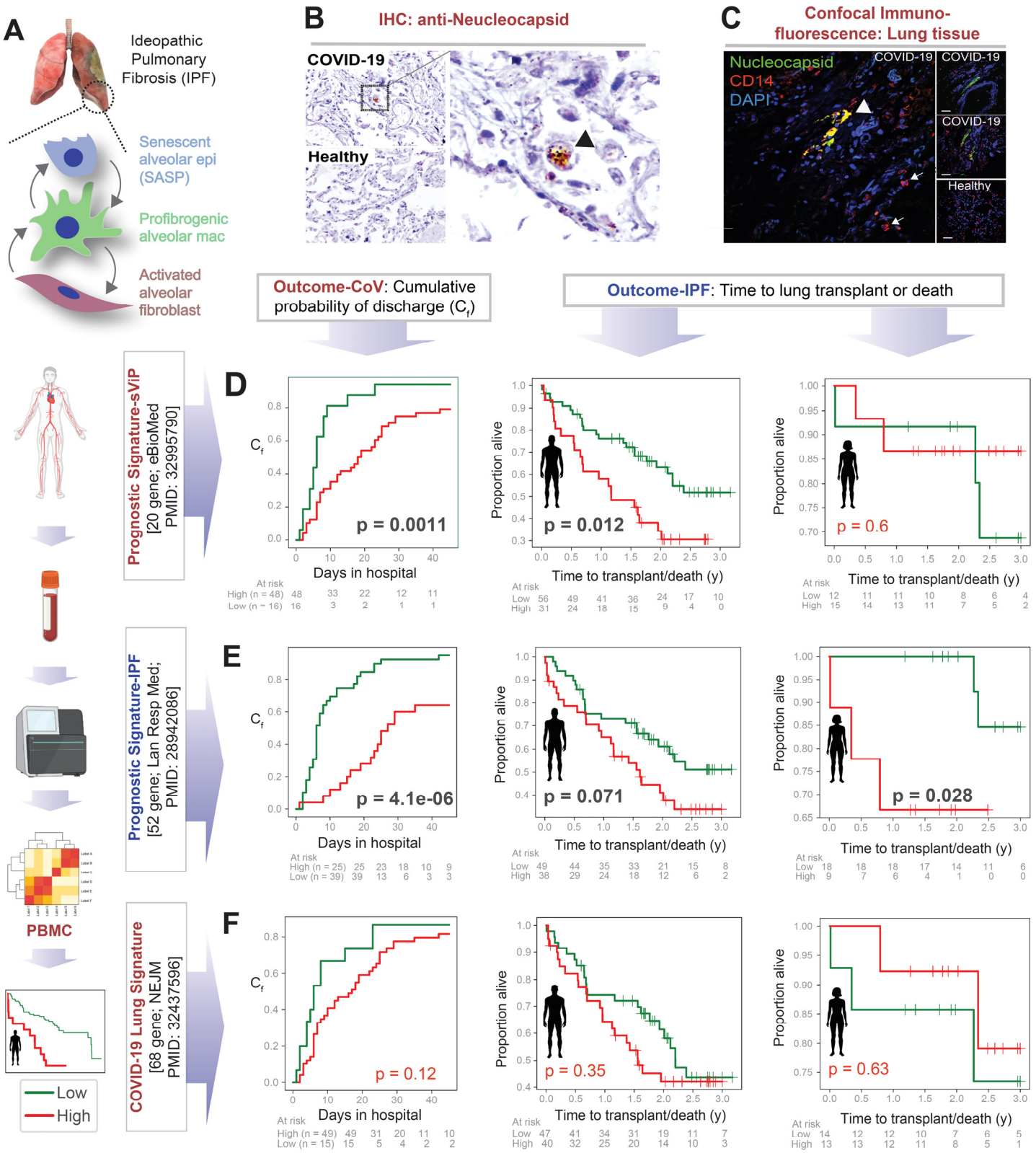

Figure 3. PBMC-derived prognostic signatures in IPF also prognosticate outcome in COVID-19.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A. Schematic showing the current understanding of the complex interplay between senescent alveolar pneumocyte (AT2), reactive alveolar macrophage, and fibroblasts in the pathogenesis of pulmonary fibrosis. **B-C**. FFPE tissues from normal lungs and from the lungs of subjects deceased due to COVID-19 were analyzed for the SARS-CoV2 viral nucleocapsid protein either by IHC (B) or by confocal immunofluorescence (C). Lungs were co-stained also with CD14 (a marker for activated alveolar macrophage[53]) in C. **D.** Kaplan-Meier plots show the stratification of patients based on the degree of induction of the sViP signature in PBMCs (low and high group; determined using the StepMiner algorithm[64] and its relationship to either the cumulative probability of discharge of COVID-19 patients (D-*left*; GSE157103; derived from the hospital-free days during a 45-day follow-up), or the time to lung transplant or death in male (D-*middle*) and female (D-*right*) IPF patients (GSE28221). Statistical significances were estimated using the log rank test. **E.** Kaplan-Meier plots on the same cohorts as in D, using a PBMC-derived 52-gene IPF-specific prognostic signature that was previously validated on numerous independent IPF cohorts[132]. **F**. Kaplan-Meier plots on the same cohorts as in D-E, using a CoV-lung derived signature.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

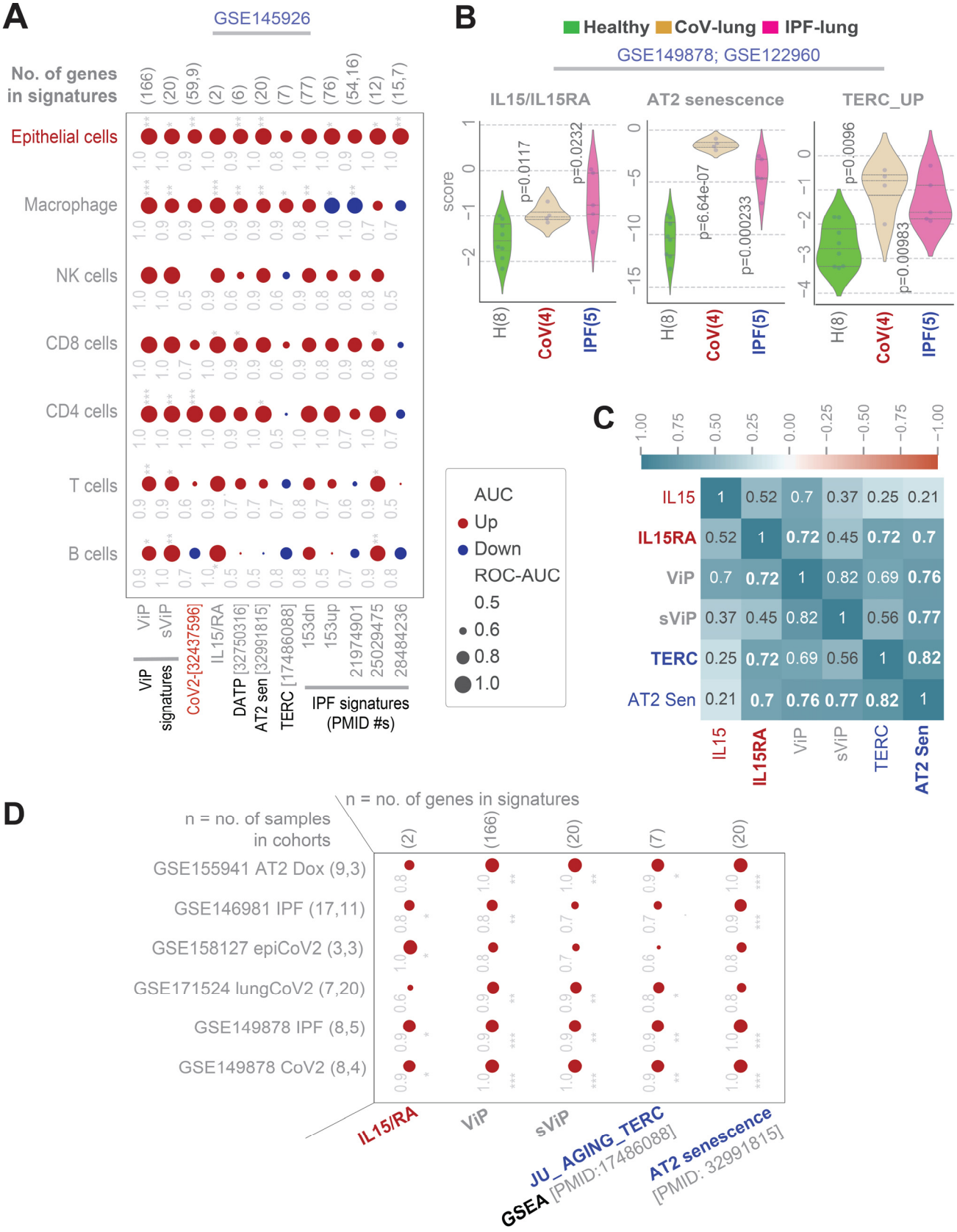

Figure 4. Identification of shared AT2 cytopathies in COVID-19 and IPF.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A. Bubble plot of ROC-AUC values (radii of circles are based on the ROC-AUC) demonstrating the direction of gene regulation (Up, red; Down, blue) for the classification of various cell types between healthy and CoV lung based on various ViP/sViP and IPF gene signatures in **Fig 2A-B** alongside *IL15/IL15RA* and additional signatures of AT2 cytopathies that are encountered and implicated in IPF (DATP, damage associated transient progenitors; Sens, senescence; TERC, *Te*lomerase *R*NA *C*omponent). Numbers in parenthesis indicate PMIDs. **B.** Violin plots display the extent of induction of a *IL15/IL15RA* composite score (B-*left*), or a signature of AT2 senescence[63] (B-*middle*), or a gene set for telomere dysfunction (JU_AGING_TERC[133]) in a pooled cohort of healthy, CoV and IPF lung samples (GSE149878; GSE122960). Welch's two sample unpaired t-test is performed on the composite gene signature score (z-score of normalized tpm count) to compute the *p* values. In multi-group setting each group is compared to the first control group and only significant p values are displayed. **C.** Correlation matrix showing the Pearson's correlation coefficient between different gene signatures in the pooled healthy, CoV and IPF lung dataset (GSE149878; GSE122960). **D.** Bubble plots of ROC-AUC values (radii of circles are based on the ROC-AUC) demonstrating the direction of gene regulation (Up, red; Down, blue) for the classification based on AT2 senescence-related signatures on other publicly available CoV and IPF bulk RNA seq datasets.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

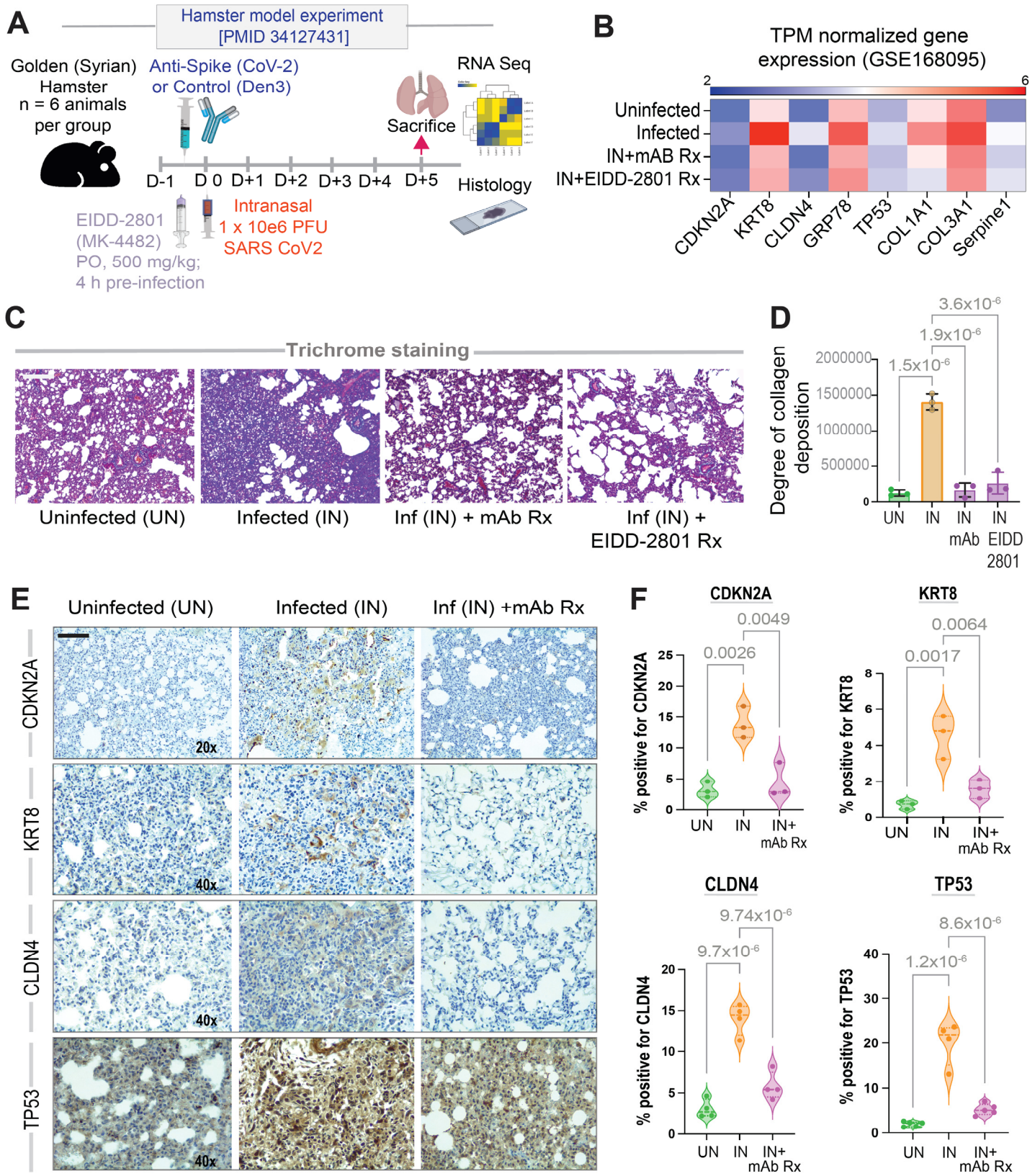

Figure 5. SARS-CoV-2-challenged hamster lungs faithfully recapitulate AT2 cytopathies in COVID-19.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A. Schematic showing the experimental design for COVID-19 modeling in golden Syrian hamsters. *Uninf*, uninfected; *Den3 and Anti-CoV-2* indicate SARS-CoV-2 challenged groups that received either a control mAb or the clone CC12.2 of anti-CoV-2 IgG[28], respectively. PFU, plaque-forming units. **B**. Heatmap showing transcripts/million (TPM)-normalized gene expression of specific genes associated with AT2 senescence (*CDKN2A*, *TP53*), damage associated transition state progenitor (DATP) arrested state (*CLDN4*, *KRT8*), ER stress (*HSPA5*/*GRP78*) and SASP-state (*SERPINE1*, *COL1A*, *COL3A*). **C-D**. Lungs harvested from the uninfected, infected, and infected but anti-CoV-2 IgG and infected but EIDD-2801 treated hamsters have been analyzed for the abundance of collagen deposition by Trichrome staining and quantified by ImageJ. Representative images are shown panel C. Scale bar = 200 µm. Results of the quantification are displayed as bar plot in D. Statistical significance was analyzed by one way ANOVA. Error bars represent S.E.M; n = 3. **E-F**. Lungs harvested from the uninfected, infected, and infected but anti-CoV-2 IgG groups were analyzed by IHC for the indicated proteins and quantified by ImageJ- IHC profiler. Representative images are shown panel E. Scale bar = 200 µm. Results of the quantification are displayed as violin plots in F. Statistical significance was analyzed by one way ANOVA; n = 4.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

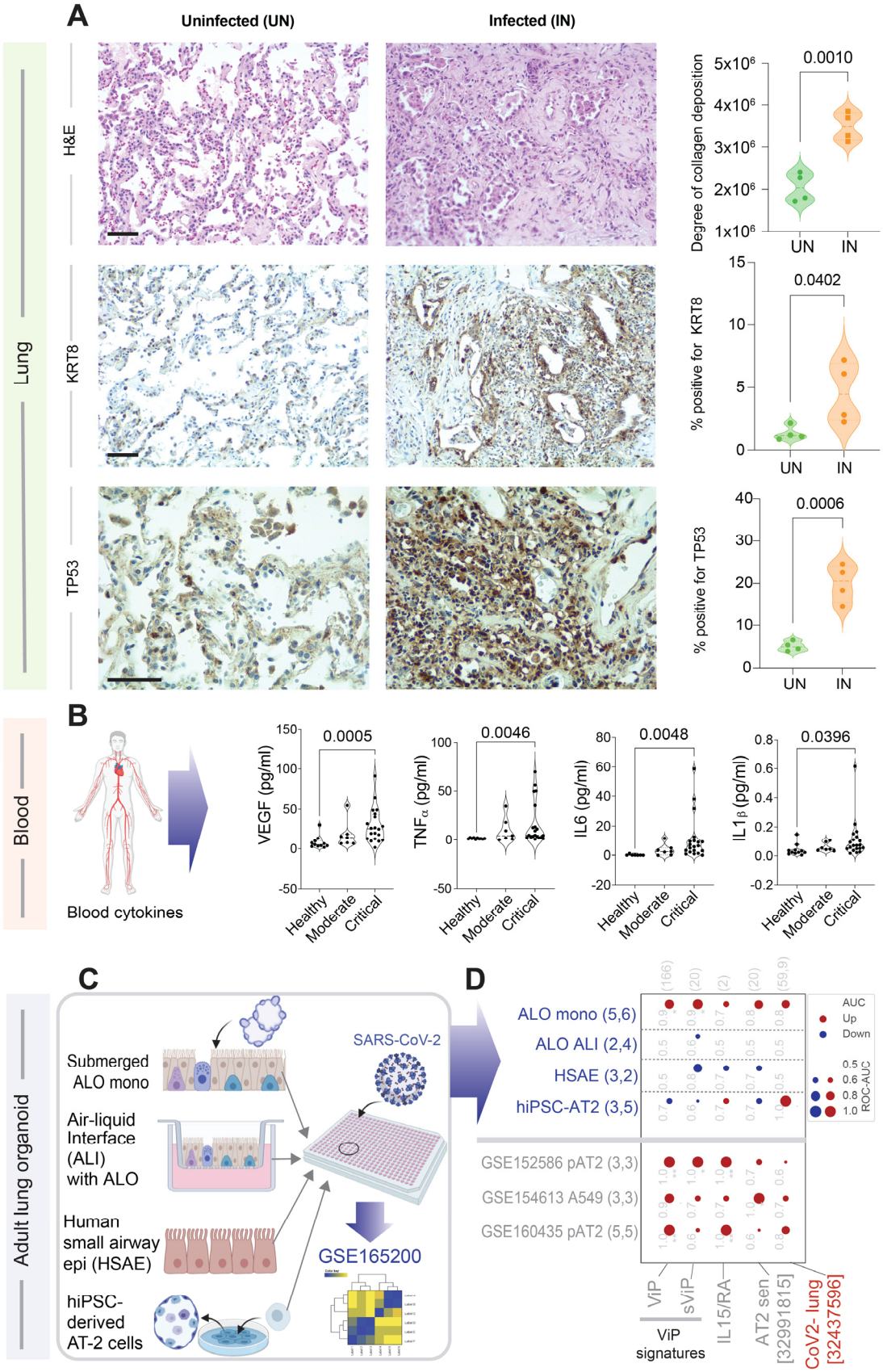

Figure 6. AT2 cytopathic features of IPF are detected in COVID-19 lung tissues and human pre-clinical organoid models of COVID-19.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A. FFPE lung tissues collected during rapid autopsy of deceased COVID-19-infected subjects and histologically normal surgically resected lung tissues were analyzed by H&E staining (*top*) and for the expression of KRT8 (a marker of DATP[60]; *middle*) and TP53 (a marker of senescence; *bottom*) by IHC. Representative fields are shown on the left. Scale bar = 200 µm. Images were quantified by IHC profiler (ImageJ) and displayed as violin plots on the right. **B**. Violin plots display the abundance of key senescence-associated cytokines in the plasma from adults with COVID-19 and healthy volunteers. Statistical significance was analyzed by unpaired t-test. See also **Supplemental Information 3-4** for patient demographics and cytokine concentrations. **C-D**. Schematic in C summarizes the various pre-clinical *in vitro* models of the human lung that were challenged with SARS-CoV-2, followed by RNA sequencing[27] (*top* rows; blue font; D). Other independent SARS-CoV-2-challenged alveolar pneumocyte datasets are listed in grey font (D). Bubble plots in D show the ROC-AUC values (radii of circles are based on the ROC-AUC) demonstrating the direction of gene regulation (Up, red; Down, blue) for the classification of uninfected vs infected samples based on the signatures (below).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

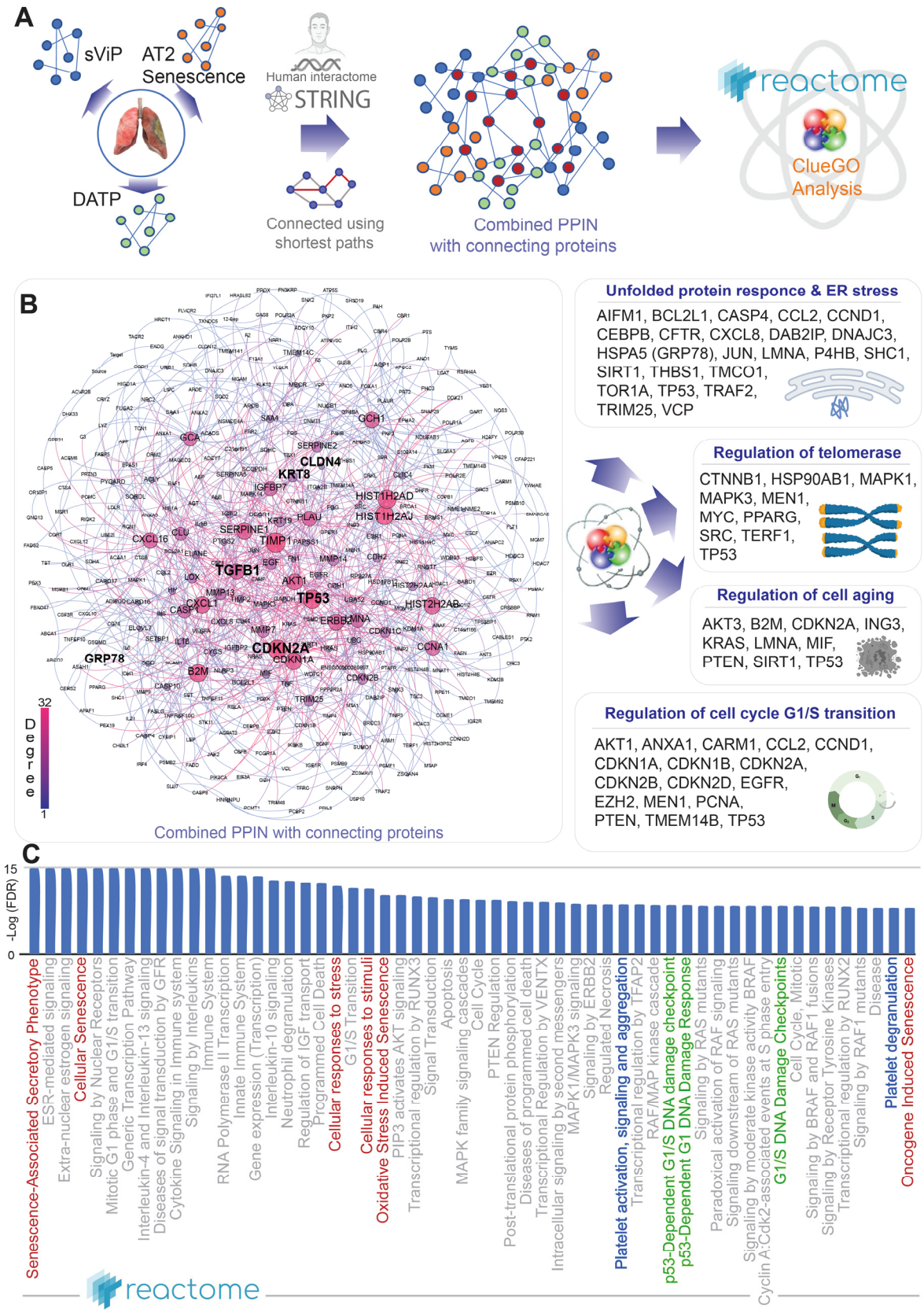

Figure 7. Protein-protein interaction network (PPIN) analysis predicts ER stress in AT2 cells as a driver pathophysiology in COVID-19

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A. Schematic summarizes the key steps in building a PPI network using as 'seeds' the genes in the sVIP, AT2 senescence and transition state arrest (DATP) signatures and connecting them using shortest paths prior to carrying out network analyses. **B**. The combined PPI network is shown on the left and the degree of connectivity is indicated. See also **Supplemental Information 5** for the complete list of nodes sorted by their degrees of connectivity. Schematics on the right summarize the results of CluGo analysis, which identifies cellular processes and key genes that may perpetuate lung fibrosis. **C**. Reactome pathway analysis of the PPI network lists the most significant pathways that are associated with the node list in the network in B. Red = senescence related processes; Green = DNA damage response pathways; Blue = Platelet activation and degranulation.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

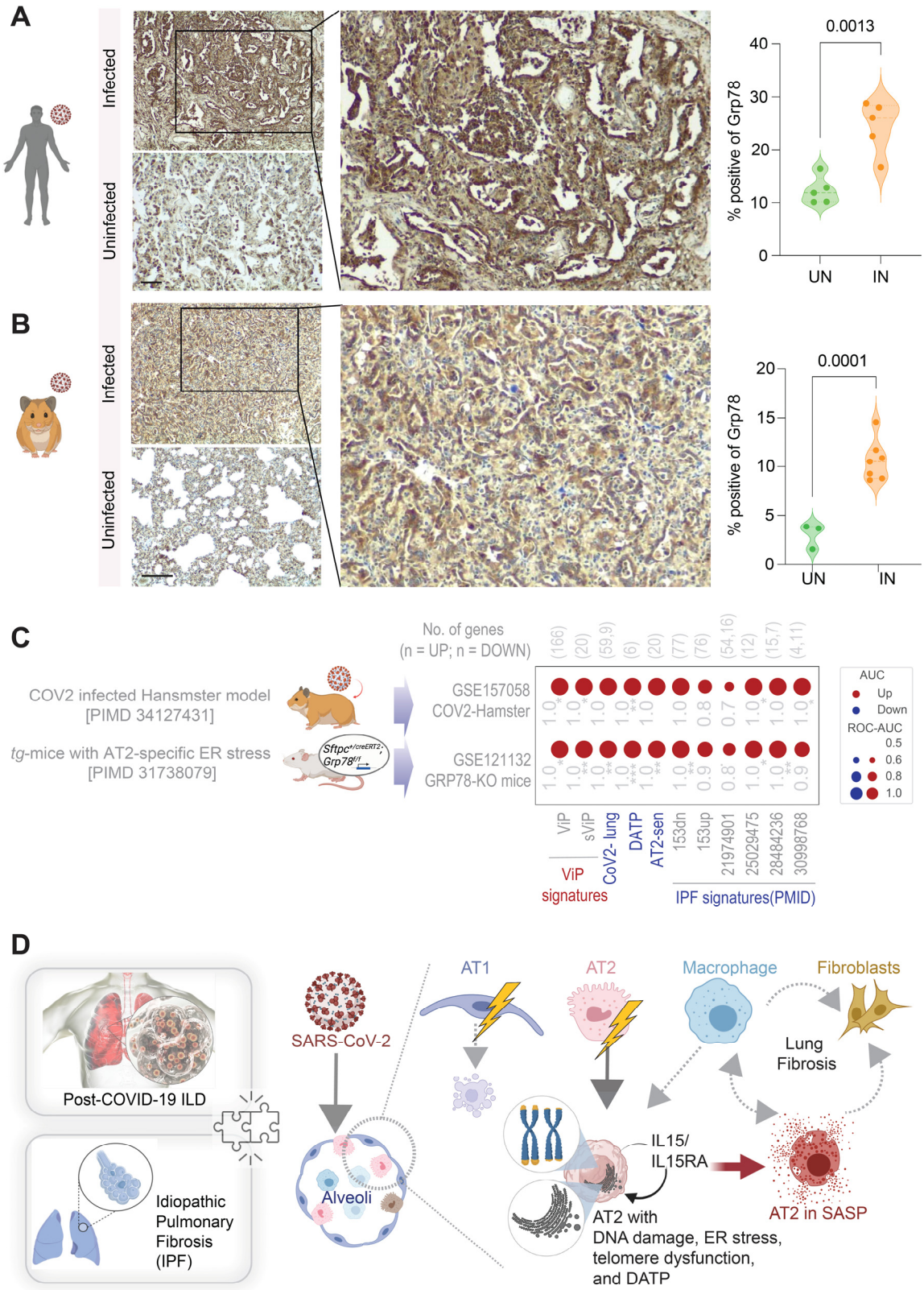

Figure 8. Induction of ER stress in AT2 cells is observed in COVID-19 and is sufficient to mimic the host immune response in COVID-19 and IPF

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.28.470269; this version posted November 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A-B. FFPE uninfected (control) and SARS-CoV-2-infected human (A) and hamster (B) lungs were analyzed for GRP78 expression by IHC. Representative fields are shown on the left. Scale bar = 200 µm. Images were quantified by IHC profiler (ImageJ) and displayed as violin plots on the right. **C**. Bubble plots show the ROC-AUC values (radii of circles are based on the ROC-AUC) demonstrating the direction of gene regulation (Up, red; Down, blue) for the classification of WT vs *GRP78*-KO murine lung (bottom row) and uninfected vs. infected hamster lung (top row) samples based on the signatures (below). **D**. Schematic summarizes the findings in this work and the proposed working model for the contributions of various alveolar cells in fueling the fibrotic progression in both IPF and COVID-19.