PNAS RESEARCH ARTICLE | IMMUNOLOGY AND INFLAMMATION OPEN ACCESS

TGFb signaling instructs a conserved fibrosis-associated cell state marked by LRRC15

Justin A. Shyer[a,1,2], Fabien Wehbe[b,1], Christopher D. Davidson[a,1], Hannah S. Bender[a], Minh Thai[a], Christian Cox[a], Alsu Missarova[a], Alexander Arlantico[a], Ben Hall[a], Ruoyu Zhang[a], David Kim[a], Anthony Altieri[c], Shakir Hasan[c], Afshin Namdar[c], Tina Chen[d], Shaheed W. Hakim[e], Cynthia Guidos[d], Hans D. Brightbill[a], Tony Kuo[b], Graham Heimberg[a], Héctor Corrada Bravo[a], Rojo A. Ratsimandresey[a], Salil Uttarwar[a], Grace Teng[a], Omar Salem[a], Mehrdad Arjomandi[f], Mark S. Wilson[a], Spyros Darmanis[a], James Ziai[a], Alexis Scherl[a], Zora Modrusan[a], Paul J. Wolters[f], Matthew B. Buechler[c,g], Jason A. Vander Heiden[a,2], and Shannon J. Turley[a,2]

Affiliations are included on p. 11.

Contributed by Shannon J. Turley; received December 17, 2025; accepted April 21, 2026; reviewed by Boris Hinz and Zemin Zhang

Fibroblasts are key potentiators of chronic disease pathophysiology. Despite their established roles in promoting pathological inflammation and tissue remodeling, activated myofibroblasts are generally characterized as a single, homogeneous cell population, obscuring critical functional distinctions. Defining the cell states, their molecular regulators, and restricted markers is critical to developing effective therapies for the treatment of fibrosis. Here, using a human lung stromal cell atlas of idiopathic pulmonary fibrosis, we identify two myofibroblast transcriptional states associated with distinct predicted biological function, regulation, and cell surface marker expression. We identify fibroblast-specific TGFb signaling as the key regulator of the mechanistic switch from a wound healing–associated and proliferative to a profibrotic myofibroblast. Further, we elucidate conserved TGFb-dependent and suppressed gene expression programs that define these states. Our findings reveal that LRRC15 is highly restricted to myofibroblasts that primarily express an extracellular matrix-remodeling gene program and illuminate that this key cell state can differentiate in the absence of an obligate inflammatory precursor intermediate. Last, we apply machine learning using a human single-cell foundation model to demonstrate broad applicability of the biology described herein to human chronic disease.

myofibroblast | fibrosis | TGFb | inflammation | wound healing

Significance

This study defines the transcriptional heterogeneity of lung myofibroblasts, revealing two key cell states associated with either inflammation or extracellular matrix (ECM) production. We identify LRRC15 as a restricted marker of the ECM production phenotype and mechanistically prove that transforming growth factor-beta (TGFb) signaling in Cdh11+ fibroblasts is required for its differentiation.

Fibrosis, a chronic pathological scarring process carried out by stromal and immune cells, accounts for approximately 30% of natural deaths worldwide (1). Idiopathic pulmonary fibrosis (IPF) is one of over 200 parenchymal pulmonary disorders, termed interstitial lung diseases (ILDs), in which fibrosis is a frequent end stage condition. IPF is chronic, progressive, and driven by the accumulation of pathologic extracellular matrix (ECM), which leads to reduced lung compliance and gas exchange, and eventual loss of organ function. The pathology of IPF is characterized by progressive replacement of the lung parenchyma with patchy, sharply demarcated peripheral rings of mature fibrosis, which are marginated by distinctive areas of active fibroplasia, forming the hallmark histologic change, termed fibroblastic foci (2). IPF is an incurable disease despite three therapeutics currently approved for treatment, Pirfenidone and Nintedanib, which have unclear mechanism(s) of action, and the recently FDA-approved PDE4B inhibitor, Nerandomilast. Critically, these medications are not universally effective and there is a strong need for novel therapies that slow the progression of disease and/or restore lung function for patients with this debilitating disease (3, 4).

Fibrosis is orchestrated by activated fibroblasts, called myofibroblasts, that produce the scar tissue responsible for organ dysfunction. Myofibroblasts contribute to disease-associated tissue remodeling via production and modification of ECM and physical tissue contraction (5). Although treatment options for patients with IPF are limited, approved and nascent therapies largely target the activation and profibrotic activity of fibroblasts, demonstrating the key role these cells play in lung fibrosis pathophysiology (3). A better understanding of the transcriptional states, regulatory signaling pathways, and unique cellular markers of myofibroblasts is key to developing improved therapies for IPF patients.

Recent efforts to characterize the fibroblast lineage using single-cell RNA sequencing (scRNA-seq) and cell atlasing approaches across species, tissues and pathologies have vastly improved our understanding of the broad organization of fibroblast transcriptional states in health and disease. Several such studies have identified shared transcriptional states of pathogenic fibroblasts present in a variety of diseases including ILDs, inflammatory bowel

Author contributions: J.A.S., H.D.B., M.B.B., J.A.V.H., and S.J.T. designed research; J.A.S., C.D.D., M.T., C.C., A. Arlantico, B.H., D.K., A. Altieri, S.H., R.A.R., G.T., M.A., M.S.W., S.D., Z.M., and P.J.W. performed research; G.H. and M.A. contributed new reagents/analytic tools; J.A.S., F.W., C.D.D., H.S.B., M.T., C.C., A.M., B.H., R.Z., D.K., A. Altieri, S.H., A.N., T.C., S.W.H., C.G., T.K., G.H., H.C.B., S.U., O.S., J.Z., A.S., and J.A.V.H. analyzed data; and J.A.S., M.B.B., J.A.V.H., and S.J.T. wrote the paper.

Reviewers: B.H., St. Michael's Hospital; and Z.Z., Peking University.

Competing interest statement: J.A.S., F.W., C.D.D., H.S.B., M.T., C.C., A.M., A. Arlantico, B.H., R.Z., D.K., H.D.B., T.K., G.H., H.C.B., R.A.R., S.U., G.T., O.S., M.S.W., S.D., J.Z., A.S., Z.M., J.A.V.H., and S.J.T. are/were employees of Genentech, Inc. S.J.T. was an employee of Genentech when the main work and revisions were carried out. S.J.T. is currently an employee of Amgen. The other authors declare no conflicts of interest.

[1]J.A.S., F.W., and C.D.D. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: shyerj@gene.com, vandej27@gene.com, or shannon.turley@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2536550123/-/DCSupplemental.

Published May 20, 2026.

disease (IBD), arthritis, and cancer with varying degrees of heterogeneity and expression of marker genes including CTHRC1, COL11A1, POSTN, ACTA2, and LRRC15 (6–8). Although these cell state markers associated with the myofibroblast phenotype have been validated across several studies, a characterization of their expression in relation to granular and functionally relevant myofibroblast activation states is lacking. Such an understanding is critical to developing novel targeted therapeutics that will have broad application in inflammatory disease.

The identification of cellular markers that faithfully associate with biological phenotypes is critical for establishing a consensus regarding activated fibroblast cell states. Indeed, a critical challenge for the field of fibroblast research is a lack of a universal nomenclature and subcharacterization of cell states. Collagen triple helix repeat containing 1 (CTHRC1) is a myofibroblast marker that has been used to identify activated myofibroblasts in the lung during fibrosis in mice and humans (9–11). On the other hand, leucine-rich repeat containing 15 (LRRC15) has been used to delineate myofibroblasts across tumor settings (8, 12, 13). Despite possible shared myofibroblast transcriptional cell states, coexpression of CTHRC1 and LRRC15 in injury and fibrosis has not been explored.

In this study, we sought to define the gene expression states, molecular regulators and restricted markers of fibrosis- and inflammation-associated fibroblasts. We find that myofibroblasts in human IPF exhibit two primary gene expression programs associated with wound healing and ECM remodeling. We define LRRC15 as a restricted marker of fibroplasia-associated gene expression and fibroblastic foci in human IPF. Using a mouse model of fibrotic lung disease, we characterize the developmental pathways and differentiation signals of these conserved gene expression programs. In vitro and in vivo studies using genetic mouse models define fibroblast TGFb signaling in Cdh11-expressing fibroblasts as the key rheostat in regulating two transcriptional states of myofibroblasts associated with inflammation and wound healing or ECM modification. These findings have important implications for therapies currently in clinical development for the treatment of fibrotic lung disease. Last, we use a newly developed single-cell foundation model assembled from 412 scRNA-seq studies, 23.4 M cells, and 30 human tissues to search for cells similar to LRRC15+ myofibroblasts and found a transcriptionally similar phenotype in a variety of inflammatory, fibrotic, and malignant pathologies.

Results

Single-Cell Atlas of Human IPF Fibroblasts Identifies Two Disease-Associated Myofibroblast Gene Expression Programs. To define IPF-associated gene expression in fibroblasts, we assembled a scRNA-seq fibroblast atlas from 76 IPF patients and 75 normal donor lungs (Fig. 1*A*) (10, 14–19). Graph-based clustering and manual annotation revealed 11 subsets of fibroblasts in normal donor and IPF lungs (Fig. 1 *A* and *B* and *SI Appendix*, Fig. S1*A*). These cell clusters are distributed across disease status, patients and data sources, and generally align with those described in previous studies that profiled human IPF fibroblasts using scRNA-seq, including subsets of adventitial fibroblasts (*PI16*+, *CD34*+; C5 and C3), lipofibroblasts (*PLIN1*+, *IL6*+; C4), alveolar fibroblasts (*NPNT*+, *TCF21*+; C1 and C2), and myofibroblasts (*POSTN*+, *CTHRC1*+; C6 and C7) (*SI Appendix*, Fig. S1 *B–H*) (10, 14, 15). Distinct from many previous studies, we annotated two clusters of myofibroblasts, characterized based on their enriched expression of genes associated with extracellular matrix production, cross-linking, and contractility and lack of expression of SMC lineage defining genes: Myofib-1 (C7) and Myofib-2 (C6) (*SI Appendix*, Fig. S1*I*). These two myofibroblast populations are distinguished by differential expression of key genes including *ADAM12 and GJB2* (*SI Appendix*, Fig. S1*J*).

We performed neighborhood-based analysis using a suite of Milo-methods and estimated both differential abundance (DA) and differential expression (DE) on a fine-grained, neighborhood level (20). Leveraging neighborhood assignment within fibroblast lineages, we approximated individual normal donor- and IPF-cell states and relied on the DA of those neighborhoods as a proxy for disease progression. Through this lens, we find that myofibroblasts exhibit the greatest association with IPF, with Myofib-1 capturing the predicted most differentiated fibroblast state along the DA axis (Fig. 1*C*). These results are suggestive of a differentiation trajectory toward the Myofib-1 state during IPF progression. We further corroborated these findings using pseudotime analyses, which reveal a predicted cell lineage ending at Myofib-1 that originates from alveolar fibroblasts and traverses through activated alveolar fibroblasts and Myofib-2 (*SI Appendix*, Fig. S1 *K* and *L*).

We next identified coregulated gene expression programs in the normal and IPF lung through a cluster-agnostic topic modeling approach using consensus non-negative matrix factorization (cNMF). This analysis identified 18 gene expression modules representing fibroblast cell identity and cell activity programs (*SI Appendix*, Fig. S2*A*). The topics identified represent gene expression programs associated with adventitial (H4, H5 and H6), alveolar (H1 and H12), and lipofibroblast (H11) cell identity as well as IPF enriched activity programs associated with proliferative (H17), activated alveolar fibroblast (H7) and myofibroblast (H2 and H15) cell activity (*SI Appendix*, Fig. S2*B*). Myofibroblast cell activity Topics H2 and H15 were selected for further analysis based on their upregulation in disease and distinct patterns of expression across the IPF fibroblast lineage (Fig. 1 *D* and *E*). Topic H2- and H15-associated gene expression was reciprocally enriched in Myofib-2 and Myofib-1, respectively (Fig. 1 *D* and *E* and *SI Appendix*, Fig. S2*B*). These data suggested that Topics H2 and H15 represent distinct, coregulated myofibroblast gene expression activation states in IPF fibroblasts. Gene ontology analysis of the top 25 genes, with genes ranked by cNMF spectra score, in Topic H2 highlighted categories associated with wound healing (*ANXA1*, *ANXA2*, *TNFRSF12A*, *TIMP1*, *CAV1*, *HIF1A*, *SERPINE1*), maintenance of location/trafficking (*THY1*, *TMSB4X, CAV1, ANXA1, CALM2, TMSB10*), and regulation of apoptosis (*LMNA, TNFRSF12A, CAV1, HIF1A, CTSC, SERPINE1*) (Fig. 1*F*). Distinctly, gene ontologies significantly enriched for the top 25 genes in Topic H15 were exclusively associated with extracellular matrix organization (*COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, POSTN, TGFBI, FAP, MMP14*) (Fig. 1*F*). Pathway analysis using the Pathway Responsive Genes for Activity Inference (PROGENy) tool identified distinct potential upstream regulators of myofibroblast associated Topics H2 and H15. Topic H2 was enriched for growth factor signaling pathways, including EGFR and VEGF, as well as signaling through PI3K (Fig. 1*G*). In contrast, Topic H15 showed strong enrichment for TGFb and WNT pathway activation (Fig. 1*G*). Additionally, we assessed the relationship between DA of cell states and median DE of the top 50 genes in topics H2 and H15. While we observe that overall patterns for H2 and H15 are similar, and both topics contain genes that are generally upregulated at later stages of disease (Fig. 1*H*), we note that topic H15 is specifically restricted to the predicted more differentiated disease cell states (defined as neighborhoods with high positive DA log fold-change), while genes from topic H2 appear to be more systematically upregulated across the

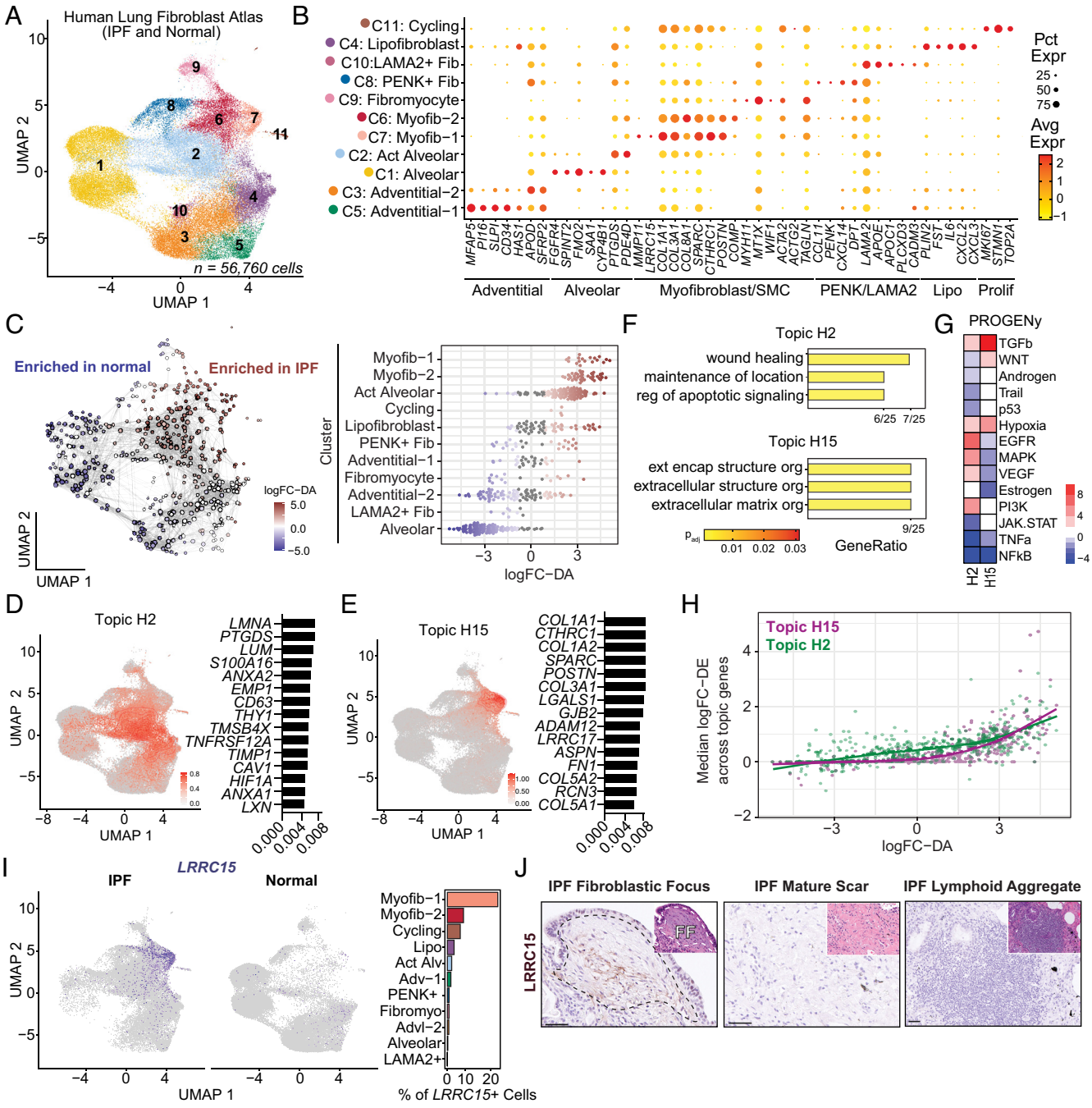

Fig. 1. Novel single-cell atlas of human IPF fibroblasts identifies disease-associated myofibroblast gene expression programs. (*A*) UMAP representation of mesenchymal clusters (color and number) from combined normal and IPF lung tissue scRNA-seq. (*B*) Dotplot of selected marker genes defining mesenchymal clusters (y-axis, dot color and number) in (*A*) showing the percent of cells within a cluster expressing each gene (dot size) and average normalized expression (dot color scale). (*C*) *Left*, UMAP representation of Milo neighborhoods (points; size represents the number of cells in each neighborhood) and estimation of neighborhood differential abundance (color scale, $\log_2$ fold-change; white is nonsignificant). *Right*, Differential abundance of Milo neighborhoods (x-axis, log scale; color scale) across cell types (y-axis) with neighborhood cell type determined by the most common cell type within a neighborhood. UMAP of mean gene spectra score across all mesenchymal cells (color scale) with barcharts of individual spectra scores (x-axis) for the top 15 genes (y-axis) for (*D*) Topic H2 and (*E*) Topic H15. (*F*) Top 3 Gene Ontology terms significantly enriched for the top 100 genes in Topic H2 (*Top*) and Topic H15 (*Bottom*). (*G*) Pathway enrichment scores (color scale) for PROGENy pathways (y-axis) across selected topics (x-axis). (*H*) Median $\log_2$ fold-change from differential expression testing (y-axis) for the top 50 genes in topics H15 (purple) and H2 (green) vs. median $\log_2$ fold-change from differential abundance testing (x-axis) across Milo neighborhoods. (*I*) *Left*, cells expressing any level of *LRRC15* transcript (blue) across all cells (gray); *Right*, the percentage of cells expressing transcript within each mesenchymal cluster (y-axis). (*J*) Representative immunohistochemistry staining for LRRC15 with *Inset* H&E in the (*Left*) FF, (*Middle*) mature scar tissue and (*Right*) a lymphoid aggregate from human IPF lung tissue.

manifold. These data suggest an IPF progression association with both Topic H15-expressing myofibroblasts and TGFβ-dependent gene programs. Together, these data demonstrate that myofibroblasts exhibiting distinctly regulated gene programs are present in the diseased IPF lung.

LRRC15 Marks Myofibroblasts in the IPF Fibroblastic Focus (FF). Leucine-rich repeat containing protein 15 (LRRC15), a transmembrane protein with documented expression in myofibroblastic cancer-associated fibroblasts, was highly enriched for expression in Myofib-1 cells and exhibited a high

spectra score for Topic H15 (Fig. 1*I* and *SI Appendix*, Fig. S2*C*). Immunohistochemistry (IHC) analysis of LRRC15 in the human IPF and normal donor lung identified LRRC15 protein expression as highly restricted to fibroblastic foci (FF), the hallmark pathological lesion in IPF, as compared to inflamed regions and fully formed scar tissue in IPF lungs, as well as normal alveolar tissue from non-IPF donor lungs (Fig. 1*J* and *SI Appendix*, Fig. S2*D*). Strikingly, all annotated FF analyzed exhibited LRRC15 expression associated with spindle-shaped cells and LRRC15 was not detected outside of the FF. These data indicate that LRRC15 marks a spatially restricted subset of fibrosis-associated fibroblasts in IPF.

To determine the spatial regulation of LRRC15+ myofibroblast gene expression, we relied on data published by Eyres et al., wherein the authors isolated regions of interest in normal and IPF lung tissue for spatial transcriptomic sequencing (20). A 96 gene signature derived to be specific to *LRRC15*+ Myofib-1 was enriched in IPF FF compared to diseased or healthy blood vessels, alveolar septae, or immune infiltrate regions (AS, BV, DAS, II) (*SI Appendix*, Fig. S2*E*), consistent with the localization of LRRC15 protein and further evidencing a clear histopathological niche for LRRC15+ myofibroblasts.

LRRC15+ Myofibroblasts Emerge in Response to Bleomycin-Induced Mouse Lung Injury. Elucidating the dynamics of LRRC15 expression in human patients presents a significant challenge, primarily due to poor access to early disease tissue sampling. Therefore, we examined LRRC15 expression in an established preclinical mouse model of lung fibrosis. Intratracheal delivery of bleomycin through passive inhalation leads to interstitial fibrosis preceded by a transient inflammatory response (21). This model elicited fibrosis that peaked by day 24 and was accompanied by an increase in fibrotic gene expression in whole lung lysates from bleomycin-treated animals compared to those that received saline (*SI Appendix*, Fig. S3 *A–C*). To enrich for live fibroblasts in flow cytometry and fluorescence activated cell sorting (FACS) following bleomycin administration, we use surface expression of the transmembrane glycoprotein podoplanin (PDPN), a marker of activated fibroblasts in tumors and other inflammatory and fibrotic settings, in epithelial and immune lineage negative cells (11, 21, 22). To characterize LRRC15 protein expression dynamics, we profiled digested single-cells from the lungs of mice at days 7, 14, and 24 following bleomycin delivery. Cell surface LRRC15 expression was detected by flow cytometry exclusively on PDPN expressing stromal cells at the fibrosis-associated timepoints day 14 and 24 and was absent in saline treated control lungs (Fig. 2 *A* and *B* and *SI Appendix*, Fig. S3 *D* and *E*). High dimension cytometry by time of flight (CyTOF) analysis confirmed the restricted expression of LRRC15 to fibrotic-timepoint associated PDPN-expressing cells (*SI Appendix*, Fig. S3 *F* and *G*). In situ analysis by immunofluorescence microscopy of lung tissue at day 24 post–bleomycin delivery confirmed the expression of LRRC15 on spindle shaped cells in regions of fibrotic remodeling as determined by positive trichrome staining in serial sections (*SI Appendix*, Fig. S3*H*). Suggestive of profibrotic function, the number of LRRC15-expressing cells per milligram of tissue in the lung 24 d following bleomycin injury correlated with total hydroxyproline (OHP), a measurement of collagen fiber deposition, and the histological fibrosis severity score (*SI Appendix*, Fig. S3*I* and Fig. 2*C*). Together, these data demonstrate that LRRC15-expressing fibroblasts emerge in mice in response to profibrotic injury in nonmalignant settings and correlate with tissue fibrosis, mirroring findings presented here from human IPF.

scRNA-seq Identifies Conserved Transcriptional States of LRRC15+ Myofibroblasts. To define the transcriptional state of LRRC15-expressing stromal cells and their precursors following alveolar injury in mice, we performed scRNA-seq of viable PDPN+ cells from steady-state lungs and lungs at days 7, 14, and 24 following intratracheal delivery of bleomycin. To ensure adequate representation of LRRC15-expressing cells, we included a separate condition of FACS sorted LRRC15-expressing cells at day 24 postinjury (*SI Appendix*, Fig. S4*A*). We identified 16 clusters of stromal cells in the steady-state and injured mouse lung (Fig. 2 *D* and *E*). The cell clusters we identified generally align with those described in previous studies that profiled mouse lung stroma using scRNA-seq (9, 10, 21, 22) (Fig. 2*E* and *SI Appendix*, Fig. S4*B*). As in the human analysis, myofibroblasts are defined based on their enriched expression of ECM, contractility, and cross-linking genes and lack of SMC lineage defining genes (*SI Appendix*, Fig. S4*C*).

Injury-associated cell subsets segregated based on the timepoint analyzed (Fig. 2 *F* and *G*). At day 7, the early inflammatory timepoint, emergent cell states derived from alveolar (C13, C16) and adventitial (C6) cells expressed previously described genes associated with tissue damage and inflammation, including serum amyloid A3 (*Saa3*) and lipocalin 2 (*Lcn2*), and interferon related genes such as interferon regulatory factor 7 (*Irf7*) and interferon-stimulated gene 15 (*Isg15*) (*SI Appendix*, Fig. S4 *D* and *E*) (9). Although inflammatory gene induction was largely conserved between the alveolar and adventitial derived subsets, some genes, including CXC motif chemokine ligand 13 (*Cxcl13*), showed lineage specific expression in adventitial derived cells (*SI Appendix*, Fig. S4*F*). Therefore, transcriptionally distinct subsets of lung fibroblasts induce shared and lineage-specific inflammatory gene programs early following alveolar injury.

At fibrosis-associated timepoints day 14 and 24, distinct cell states emerge with retained adventitial (C6), smooth-muscle like (C7), and alveolar (C5, C9, C10, and C3) phenotypes (Fig. 2 *F* and *G*). Collagen triple helix repeat containing 1 (*Cthrc1*), a broad marker of profibrotic fibroblasts in this model, was expressed by cells descendant of all three lineages, with the greatest proportion of the cells derived from alveolar fibroblasts (*SI Appendix*, Fig. S4*G*). These data are in agreement with previous reports predicting a multilineage fibroblast response to induced alveolar injury, with the majority of myofibroblasts derived from alveolar fibroblasts (9, 22).

Consistent with surface protein detection by flow cytometry and human data from IPF patients, *Lrrc15* mRNA was measured only in injured conditions and was enriched at fibrosis-associated timepoints day 14 and day 24 (*SI Appendix*, Fig. S4 *H* and *I*). *Lrrc15* expression was enriched in the Myofib-2 cluster (C3) with expression also detected in Myofib-1 (C10) as well as injury-associated peribronchiolar (C4), SMC-like (C7 and C12), and proliferative (C11) cells (*SI Appendix*, Fig. S4*J*). Notably, *Lrrc15* expression was not detected in emergent cells descendant of adventitial cells (C6) or in nonmyofibroblast populations of activated alveolar fibroblasts (C5, C9). These data support a model wherein LRRC15 expression in the bleomycin lung fibrosis model is restricted to alveolar fibroblast- and SMC- derived myofibroblasts.

To determine if the gene expression programs we identified in human IPF were conserved and could be mechanistically investigated using this model, we again performed cNMF. This analysis identified 17 gene expression modules representing cell identity and cell activity programs (*SI Appendix*, Fig. S5*A*). The top represented 13 topics described the relevant biological phenotypes

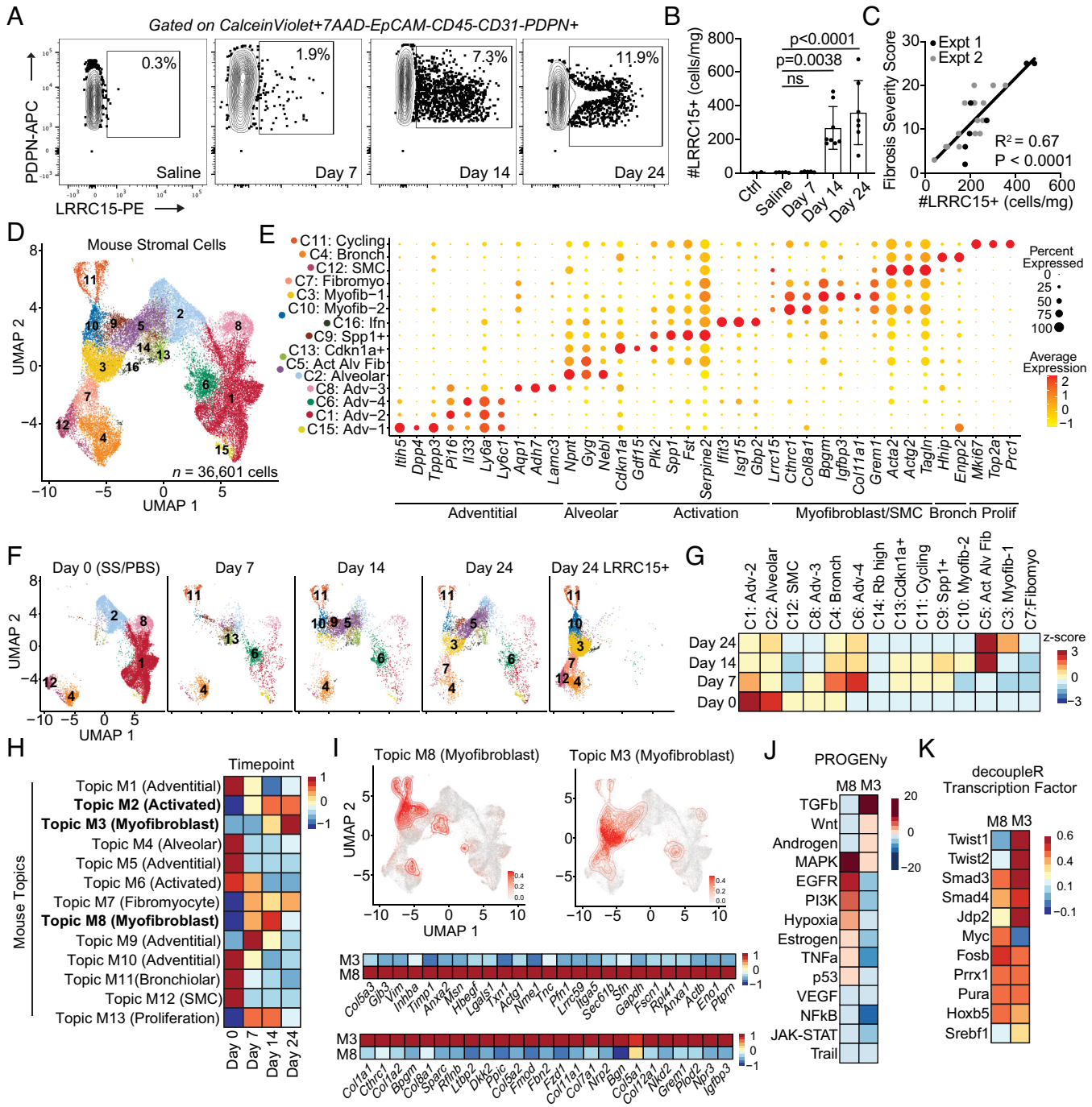

Fig. 2. Single-cell RNAseq identifies transcriptional states of LRRC15+ myofibroblasts and gene expression programs induced in injury-associated stromal cell populations in mice. (*A*) Representative flow cytometry plots and (*B*) quantification of the number of LRRC15+ cells per mg of tissue (n = 2-8 per group). (*C*) Correlation of the number of LRRC15+ cells per mg of tissue with histological fibrosis severity score (data from two experiments combined; n = 24). (*D*) UMAP representation of mesenchymal clusters (color and number) from control and bleomycin treated mouse lung scRNA-seq, composed of isolated PDPN+ cells from days 0 (control), 7, 14, and 24 and isolated LRRC15+ cells from day 24 following bleomycin treatment. (*E*) Dotplot of selected marker genes defining mesenchymal clusters (y-axis, dot color and number) in (*D*) showing the percent of cells within a cluster expressing each gene (dot size) and average normalized expression (dot color scale). (*F*) UMAP representation of cells and clusters in (*C*) separated by time point and isolation marker. (*G*) Row scaled heatmap showing the percentage of cells (color scale) within each cluster (x-axis) across time points (y-axis). (*H*) Heatmap of gene signature score for the top 50 genes of each transcriptional topic by time point. (*I*) (*Top*) UMAP of gene signature score across all mesenchymal cells (points, color scale) overlaid with density contours (lines) for Topic M8 (*Left*) and Topic M3 (*Right*). (*Bottom*) Gene spectra scores (color scale) for the top 25 genes in Topic M8 (*Top*), Topic M3 (*Bottom*). (*J*) Pathway enrichment scores (color scale) for PROGENy pathways (x-axis) across selected topics (y-axis). (*K*) Transcription factor (x-axis) target gene set enrichment scores (color scale) for selected topics (y-axis). Data in (*B*) and (*C*) are representative of two independent experiments and are mean ± SD. Statistics were calculated using two-tailed, unpaired Student's t test. The solid line in (*C*) represents simple linear regression of all data points.

present in the dataset (Fig. 2*H* and *SI Appendix*, Fig. S5 *B* and *C*). To assess conservation, we constructed a mapping between human and mouse topics by calculating the Jaccard index, a measure of set overlap, across the top 100 gene orthologs for all topic pairs. These analyses determined that the LRRC15-associated Topic H15 was most similar to mouse Topic M3, while Topic H2 was most closely related to Topic M8 (*SI Appendix*, Fig. S5*D*). Collectively, our analysis reveals shared myofibroblast cell states in mouse and human where M3/H15 (dominant LRRC15-expressing cells) and M8/H2 share high transcriptional similarity.

Additionally, gene ontology, upstream regulator and transcription factor prediction analyses revealed conserved biologies associated with Topics M8 and M3. Topic M8- and M3-associated gene expression was enriched in myofibroblasts at early day 14 and late day 24 fibrotic timepoints, respectively (Fig. 2 *H* and *I*). Gene ontology analysis of the top genes in topic M8 highlighted categories associated with wound healing (*Anxa1*, *Anxa2*, *Actg1*, *Hbegf*, *Timp1*), cell growth (*Inhba*, *Hbegf*, *Tnc*, *Sfn*, *Eno1*), and leukocyte cell adhesion (*Itga5*, *Msn, Lgals1*, *Anxa1*, *Actb*) (*SI Appendix*, Fig. S5*E*). Gene ontology analysis of the top genes in topic M3 highlighted categories associated with extracellular matrix modification (*Col1, Col5, Col8, Col11, Col12*) and canonical Wnt signaling (*Cthrc1*, *Dkk2*, *Fzd1*, *Nkd2*, *Grem1*) (*SI Appendix*, Fig. S5*F*). Collectively, these data highlight the temporally regulated functions of myofibroblasts in response to injury, coordinating immune system activation and wound closure early in response to damage before switching to tissue repair and resolution programs.

Reciprocal enrichment of topics M3 and M8 was suggestive of distinct signaling pathways regulating and potentially cross-repressing these two gene programs. The PROGENy pathway activity inference tool identified potential upstream regulators of injury-induced topics (Fig. 2*J*). Topic M8 enriched for signaling networks associated with inflammation, including TNFa and MAPK signaling, hypoxia, and proliferation-associated category p53, consistent with a role for this gene program in inflammation and wound healing. Consistent with our human analyses, the top predicted regulators of topic M3-associated genes were TGFb and Wnt (Fig. 2*J*). Transcription factor activity inference analysis provided consistent results. Topic M8 enriched for transcription factors associated with inflammatory signaling, including AP-1 family member Fosb and TNFa signaling regulator Prrx1, and proliferation, including master regulators of proliferation Myc and Pura (Fig. 2*K*). Top genes in topic M3 were predicted to be regulated by TGFb-associated transcription factors Smad3 and Smad4 as well as the Wnt-inducible transcription factors Twist1 and Twist2 (Fig. 2*K*). *Twist1,* though not *Twist2*, mRNA was specifically enriched in the topic M3-expressing Myofib-2 cluster, further implicating its potential role in regulating gene expression in fibrosis-associated myofibroblasts (*SI Appendix*, Fig. S6 *A* and *B*).

To further characterize the differentiation of LRRC15+ myofibroblasts following lung injury, we performed trajectory inference analysis. As the alveolar fibroblast lineage is the primary source of injury-derived inflammatory and fibrogenic cells in the bleomycin lung injury model (9), we focused our analysis on only this subset of our data. Graph-based clustering of alveolar lineage cells produced 6 clusters consistent with our previous analyses (*SI Appendix*, Fig. S7 *A* and *B*). Slingshot cell lineage and pseudotime inference analyses identified three lineage trajectories all descendant from alveolar fibroblasts, two of which corresponded to myofibroblast differentiation trajectories (lineages 2 and 3; *SI Appendix*, Fig. S7 *C–F*) (23). Lineage 2, which more highly corresponded to cell states present at day 14, indicated a differentiation trajectory through the *Spp1*+ inflammatory and Myofib-2 states. Analysis of significantly differentially expressed genes across lineage 2 that overlap with the previously described gene expression programs indicated a cell state transition from Topic M2- to M8- to M3-associated gene expression (*SI Appendix*, Fig. S7*G*). This lineage favored a model wherein at early fibrotic timepoints myofibroblast differentiation progresses through an inflammatory intermediate state, in line with those described in several recently published reports in different contexts (9, 23, 24). In contrast, lineage 3, which corresponded to cell states enriched at day 24 post–bleomycin injury, was defined by variable genes overlapping with Topic M3 gene expression, suggesting that differentiation of LRRC15-expressing myofibroblasts may occur at late fibrotic timepoints without an obligate inflammatory precursor (*SI Appendix*, Fig. S7 *F* and *G*).

TGFbR2 Signaling in *Cdh11*-Expressing Fibroblasts Is Required for Lung Fibrosis Following Injury. To demonstrate a functional role of TGFbR2 signaling in regulating the myofibroblast gene expression programs we identified, we generated a novel genetically modified mouse model to delete the receptor in a relevant responsive fibroblast lineage. Cadherin-11 (CDH11), a critical cell-adhesion protein, is highly expressed in normal alveolar and fibrosis-associated activated fibroblasts in both humans and mice (*SI Appendix*, Fig. S8 *A* and *B*). Protein staining confirms broad expression in human IPF across histopathological niches including fibroblastic foci, tissue containing lymphoid aggregates and fully formed scar tissue (*SI Appendix*, Fig. S8*C*). Previous studies have documented a role for CDH11 protein in regulating fibrosis and inflammation in various contexts, including malignancy, rheumatoid arthritis, and fibrosis of the heart, liver, and lung (25–30). In lung fibrosis, CDH11 expressed on fibroblasts and macrophages promotes attachment of these cells and sustains TGFb-mediated profibrotic niches (31).

We hypothesized that CDH11-expressing alveolar and early activated fibroblasts are the precursors to profibrogenic myofibroblasts and require TGFbR2 signaling to upregulate the ECM gene program we identified and drive fibrosis. To test this hypothesis, we generated a $Cdh11^{IresCreERT2}$ mouse by inserting a tamoxifen-dependent IresCreERT2 cassette at the 3' end of the stop codon of *Cdh11*. We crossed this strain to the $Rosa26^{LSLYFP}$ to irreversibly mark *Cdh11*-expressing cells and their progeny with yellow fluorescent protein (YFP) following tamoxifen administration. Daily intraperitoneal injection of tamoxifen for 5 d in steady-state mice resulted in active cre recombinase activity in stromal cells in various tissues, including the lung, pancreas, bone, skin, liver, and heart; the highest proportion of YFP-expressing stroma was detected in the lung (*SI Appendix*, Fig. S8*D*). YFP was detected in approximately 60% of steady-state lung fibroblasts (*SI Appendix*, Fig. S8*E*). Specifically, YFP expression was highest in PDPN+ fibroblasts that are negative for both Ly6C and SCA-1, a population that is largely composed of alveolar and other lung-specialized fibroblasts (*SI Appendix*, Fig. S8*F*) (6). scRNA-seq of YFP-positive and -negative lung stromal cells revealed that this mouse model exhibits cre activity in alveolar fibroblasts and lung-specialized adventitial fibroblasts, with minimal activity in less specialized fibroblasts (Adventitial-1), peribronchial cells, SMCs, or pericytes (*SI Appendix*, Fig. S8 *G* and *H*).

To first confirm that *Cdh11*-expressing lung fibroblasts are a major precursor to profibrogenic myofibroblasts in the bleomycin model, we performed a lineage tracing study to track prelabeled YFP fibroblasts in the lung after injury. At day 24 post–bleomycin injury, approximately 60% of LRRC15-expressing myofibroblasts in the fibrotic lung were YFP+ and thus descendent from *Cdh11*-expressing steady state cells (*SI Appendix*, Fig. S8*I*). As *Cdh11* is expressed in alveolar fibroblasts, this result is in agreement with previous reports in the bleomycin model (9). Notably, in this mouse model, no endothelial, immune, or epithelial lineage cells become labeled as YFP+ following bleomycin injury (*SI Appendix*, Fig. S8*J*). As previous studies have identified CDH11-positive immune and epithelial cells in mouse lungs, it is important to note that this may be a feature of the genetic mouse model and tamoxifen regimen used (26, 31).

We next examined whether sustained signaling through TGFbR2 in *Cdh11*-expressing alveolar fibroblasts and activated fibroblasts was

required for myofibroblast activation and subsequent fibrosis after bleomycin lung injury using a *Cdh11*^{IresCreERT2}*Tgfbr2*^{flox}*Rosa26*^{LSLYFP} mouse model. To assay the activation of LRRC15-expressing myofibroblasts, we performed flow cytometry analysis on cells isolated from the lungs of these mice on day 24 post–bleomycin dosing. Strikingly, there was an almost complete loss of LRRC15-expressing cells across the whole lung (Fig. 3*A*). Concurrent with this global loss of LRRC15+ myofibroblasts, overall tissue fibrosis was completely abolished in *Cdh11*^{IresCreERT2}*Tgfbr2*$^{fl/fl}$ mice compared to *Cdh11*^{IresCreERT2}*Tgfbr2*$^{wt/wt}$ control mice, with no observed accumulation of collagen by trichrome staining (Fig. 3*B*). Blinded pathologic scoring of fibrosis confirmed this significant reduction in fibrosis, with lungs from *Cdh11*^{IresCreERT2}*Tgfbr2*$^{fl/fl}$ mice scoring similar to healthy lung and significantly lower than lungs from *Cdh11*^{IresCreERT2}*Tgfbr2*$^{wt/wt}$ mice (Fig. 3*C*). Furthermore, qPCR of whole lung lysate showed a significant reduction in expression of the fibrosis-associated ECM gene *Col3a1* (Fig. 3*D*). Given the role of the immune system in driving inflammation and fibrosis after injury, we assayed the immune cell composition by flow cytometry. Alongside the reduction in myofibroblast numbers and deposition of fibrotic ECM, macrophages and CD4+ T cells were also significantly reduced in the lungs of *Cdh11*^{IresCreERT2}*Tgfbr2*$^{fl/fl}$ mice (Fig. 3 *E* and *F*). Reciprocally, neutrophils and monocytes were significantly increased in the lung parenchyma of *Cdh11*^{IresCreERT2}*Tgfbr2*$^{fl/fl}$ mice (Fig. 3 *G* and *H*). Strikingly, sorted fibroblasts from *Cdh11*^{IresCreERT2}*Tgfbr2*$^{fl/fl}$ mice exhibited a marked reduction in chemokine signaling, suggesting that TGFb signaling in fibroblasts may directly regulate immune cell recruitment into the fibrotic lung (Fig. 3*I*). These data demonstrate that TGFb signaling in alveolar fibroblasts and activated fibroblasts plays an essential role in regulating both inflammation and fibrosis after injury.

TGFb Signaling Is Required for LRRC15-Associated ECM Gene Expression. Our in silico analyses implicate TGFb signaling and synergy with other master regulators of inflammation, wound healing, and development as dominant regulators of injury responsive fibroblast gene expression and LRRC15+ myofibroblast development. Our experiments in *Cdh11*^{IresCreERT2}*Tgfbr2*$^{fl/fl}$ mice,

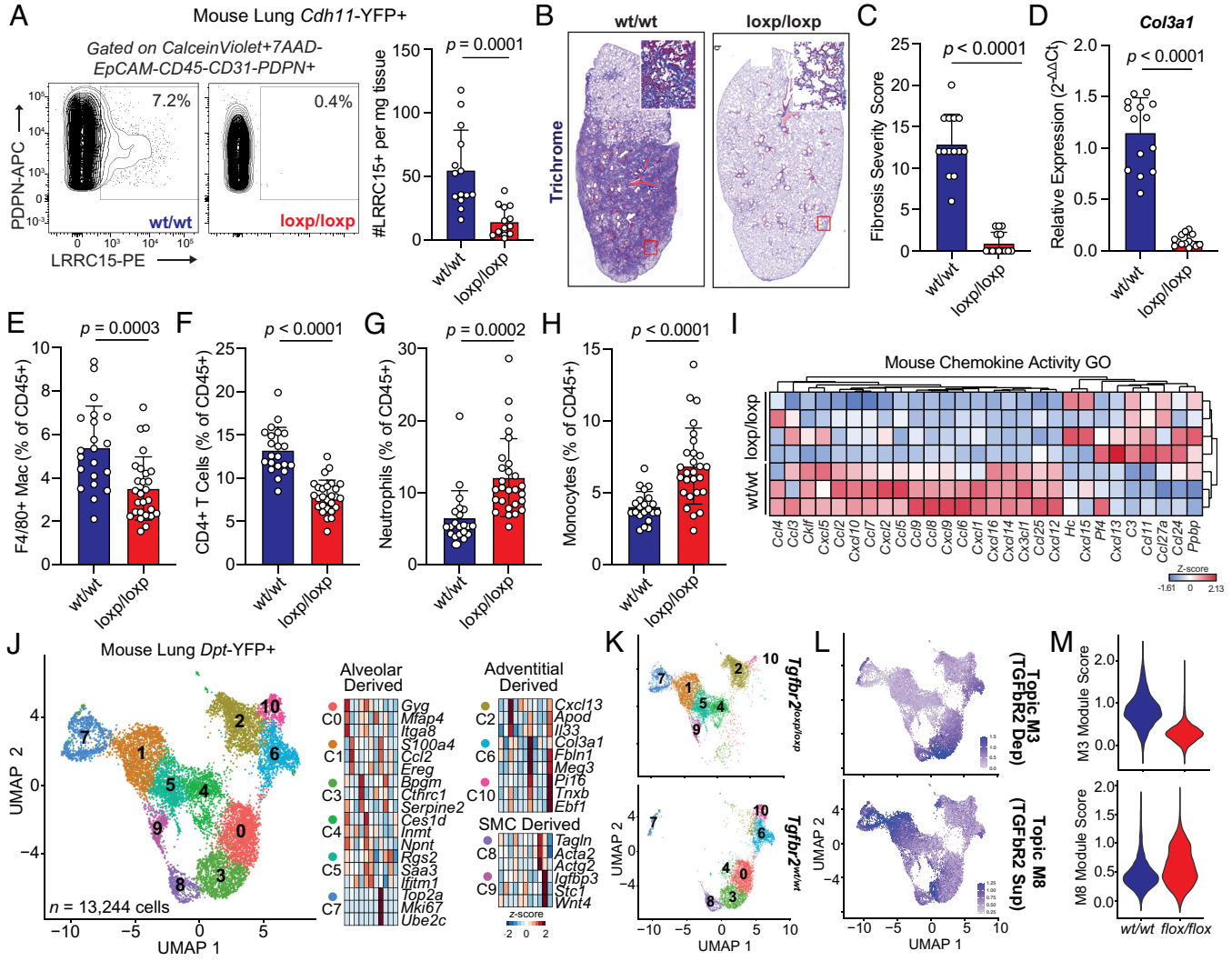

Fig. 3. TGFbR2 signaling in *Cdh11*-expressing fibroblasts is required for LRRC15+ myofibroblast differentiation and lung fibrosis and following injury. (*A*) *Left*, representative flow cytometry plots and *Right*, quantification of the number of LRRC15+ cells per mg of tissue (n = 14 per group). (*B*) Representative Mason's Trichrome staining of lung sections and (*C*) quantification of histological fibrosis severity score. (*D*) qPCR quantification of *Col3a1* in whole lung tissue lysate. Flow cytometry quantification of the proportion of (*E*) F4/80+ macrophages, (*F*) CD4+ T cells, (*G*) Neutrophils and (*H*) Monocytes of live CD45+ cells in digested lung tissue. (*I*) Heatmap of expression of mouse chemokine genes from bulk-RNA sequencing of whole lung tissue lysate. (*J*) *Left*, UMAP representation of 12,244 single PDPN+YFP+ stromal cells and right, relative average expression of indicated marker genes across clusters. (*K*) UMAP as in (*J*) split by genotype for the flox allele at the *Tgfbr2* locus. (*L*) UMAP as in (*J*) colored by topic expression and (*M*) violin plot of topic expression split by genotype for the flox allele at the *Tgfbr2* locus for (*Top*) Topic M3 and (*Bottom*) Topic M8. Data in (*A*) and (*C*–*H*) are representative of two independent experiments and are mean ± SD. Statistics in (*A*) and (*C*–*H*) were calculated using two-tailed, unpaired Student's *t* test.

combined with previous reports using alternative mouse models, have demonstrated a role for TGFb signaling in myofibroblast differentiation and tissue fibrosis following lung injury (9, 32–34). However, these studies and experiments were unable to distinguish TGFb receptor mediated effects from those due to cell-extrinsic tissue damage that persists and alterations to tissue level mechanics due to the absence of myofibroblast-mediated repair in conditional knockout mice. For example, a recently published report found that conditionally deleting *Tgfbr2* in alveolar fibroblasts prior to bleomycin injury led to decreased fibrosis but concurrent increases in tissue inflammation and animal mortality (9). To isolate the fibroblast-specific transcriptomic effects of TGFb receptor signaling in fibroblasts following injury, we used a $Dpt^{CreERT2}Tgfbr2^{flox}Rosa26^{LSL.YFP}$ mouse model to conditionally delete the receptor and lineage trace cre activity in only a small fraction of injury responsive lung fibroblasts. This strategy allowed us to define the specific gene expression programs dependent on and repressed by TGFb receptor mediated signaling in the context of a progressing fibrotic response. Notably in these experiments, we utilize this mouse model as a tool to manipulate and capture a small proportion of responsive fibroblasts to analyze transcriptomic effects of Tgfbr2 signaling in fibroblasts, not to examine the lineage relationship underpinning myofibroblast development, and do not make any conclusions regarding the lineage relationship of the cells.

Daily intraperitoneal injection of tamoxifen for 5 d resulted in active cre recombinase activity in, on average, only 25% of lung fibroblasts (*SI Appendix*, Fig. S9*A*). A proportion of all subsets of lung fibroblasts (from 20 to 40%), delineated based on expression of previously defined markers Ly6C and SCA-1, were marked by YFP expression (*SI Appendix*, Fig. S9*B*) (6). TGFbR2 depletion using this strategy does not alter overall tissue fibrosis at day 24 following bleomycin-induced injury, due to the small percentage of alveolar fibroblasts that express the cre-driving gene *Dpt*.

To assay the requirement for TGFbR2 signaling for LRRC15-expressing fibroblasts, we performed flow cytometry analysis on cells isolated from the lungs of $Dpt^{CreERT2}Tgfbr2^{flox}Rosa26^{LSL.YFP}$ mice 24 d after intratracheal delivery of bleomycin. LRRC15 expression and the number of LRRC15-expressing cells was completely diminished in YFP+ cells from $Dpt^{CreERT2}Tgfbr2^{fl/fl}$ mice compared to $Dpt^{CreERT2}Tgfbr2^{wt/wt}$ controls (*SI Appendix*, Fig. S9*C*). These data warranted a more comprehensive analysis of the transcriptional phenotype of these cells.

To address this, we performed scRNA-seq on FACS sorted YFP-expressing cells at day 24 post–bleomycin injury from $Dpt^{CreERT2}Tgfbr2^{fl/fl}$ and $Dpt^{CreERT2}Tgfbr2^{wt/wt}$ mice. After graph-based clustering, 10 groups of cells were identified, each representative of cells from multiple animals (Fig. 3*J* and *SI Appendix*, Fig. S9*D*). Cell clusters were grouped based on marker gene expression as alveolar fibroblast, adventitial fibroblast, or SMC derived (*SI Appendix*, Fig. S9*E*). The presence of multiple lineages in the dataset is consistent with our previous characterization of the $Dpt^{CreERT2}$ mouse model (*SI Appendix*, Fig. S9*B*). Strikingly, the transcriptional states of cells from $Dpt^{CreERT2}Tgfbr2^{fl/fl}$ and $Dpt^{CreERT2}Tgfbr2^{wt/wt}$ mice were completely distinct in the UMAP space, identifying TGFbR2 signaling as a key master regulator of stromal cell identity in the injured and fibrotic lung (Fig. 3*K*). In line with flow cytometry studies, *Lrrc15* mRNA expression is completely diminished in the YFP+ cells from the $Dpt^{CreERT2}Tgfbr2^{fl/fl}$ condition (*SI Appendix*, Fig. S9*F*). Other established markers of myofibroblast identity, including *Cthrc1*, *Acta2*, *Col1a1*, and *Postn* were also significantly decreased in this condition (*SI Appendix*, Fig. S9*G*).

These results led us to investigate the role of TGFbR2 signaling in regulating the injury and fibrosis gene expression programs we identified (Fig. 2). Topic M3 gene expression was completely diminished in YFP+ fibroblasts from $Dpt^{CreERT2}Tgfbr2^{fl/fl}$ mice compared to controls, demonstrating that fibroblast-specific TGFbR2 signaling is required for expression of Topic M3 genes in vivo (Fig. 3 *L* and *M*). In contrast, Topic M8 gene expression was more highly expressed, and a greater proportion of the cells were annotated as proliferating in $Dpt^{CreERT2}Tgfbr2^{fl/fl}$ mice compared to controls, indicating that TGFbR2 signaling is required to diminish inflammatory and proliferative gene programs at late fibrotic timepoints (Fig. 3 *L* and *M* and *SI Appendix*, Fig. S9 *H* and *I*).

TGFb Signaling Is Sufficient for LRRC15 and Topic H15 Gene Expression in Human Lung Fibroblasts. To determine if TGFb signaling is sufficient to induce LRRC15 and the associated myofibroblast H15 gene expression program in human cells, we employed human in vitro culture systems. TGFb1 stimulation of normal human lung fibroblasts (NHLFs) resulted in a marked increase in *LRRC15* mRNA and LRRC15 surface protein expression (Fig. 4 *A* and *B*). In human precision cut lung slices (huPCLS), which better approximate the lung microenvironment in which fibroblasts reside, TGFb1 stimulation alone was sufficient to induce *LRRC15* and all but one of the top 15 genes in human myofibroblast topic H15 (Fig. 4*C*). These data collectively demonstrate the translatability of our findings in mouse models and the critical function of TGFb to promote disease associated gene expression in the absence of other signals.

LRRC15+ Myofibroblasts Feature in Multiple Fibrotic and Malignant Diseases. Given the association of *LRRC15*+ myofibroblasts with IPF, shown here, and previous work showing an association with pancreatic ductal adenocarcinoma (PDAC) and other malignancies (12, 13, 35), we sought to determine the extent to which this cellular program can be observed across different human tissue and disease contexts. To perform this investigation, we used the SCimilarity package and accompanying v1.1 human model, which provided algorithms to retrieve similar cells to a query cell profile across a corpus of 23.4 M cells spanning 412 public sc/snRNA-seq studies (36). We constructed a consensus cellular profile from the centroid of all *LRRC15*+ cells (≥1 UMI) in cluster Myofib-1 from our human fibroblast atlas and searched across all cells within a maximum distance of 0.5 that were classified as either fibroblasts or myofibroblasts in the SCimilarity corpus, selecting the top approximately 1% of cells (SCimilarity score ≥15) as similar to our *LRRC15*+ myofibroblasts query from a total of ~2.1 M cells across 2,429 samples and 280 studies (*SI Appendix*, Fig. S10*A*). We then classified the abundance of similar cells as a percentage of the total number of fibroblast lineage cells in a study or sample, restricted to studies with ≥50 or samples with ≥50 total fibroblast lineage cells, respectively.

Cells similar to *LRRC15*+ myofibroblasts were enriched in fibrotic diseases and malignancies across multiple tissues, including lung, skin, pancreas, intestine, adipose tissue, and female reproductive organs (Fig. 4*D* and *SI Appendix*, Fig. S10*B*). In particular, we saw a high mean percentage of cells similar to *LRRC15*+ myofibroblasts compared to healthy samples (0.5%) in i) the fibrotic diseases of IPF and other ILDs (16.1%), scleroderma (systemic sclerosis, SSc) (22.6%), hidradenitis suppurativa (10.2%), nonalcoholic fatty liver disease (NAFLD) (8.0%), lymphangioleiomyomatosis (LAM) (0.9%), and eczema (atopic dermatitis) (0.6%), and ii) cancer (8.1%) including PDAC, ovarian serous carcinoma, hepatocellular carcinoma, and intestinal

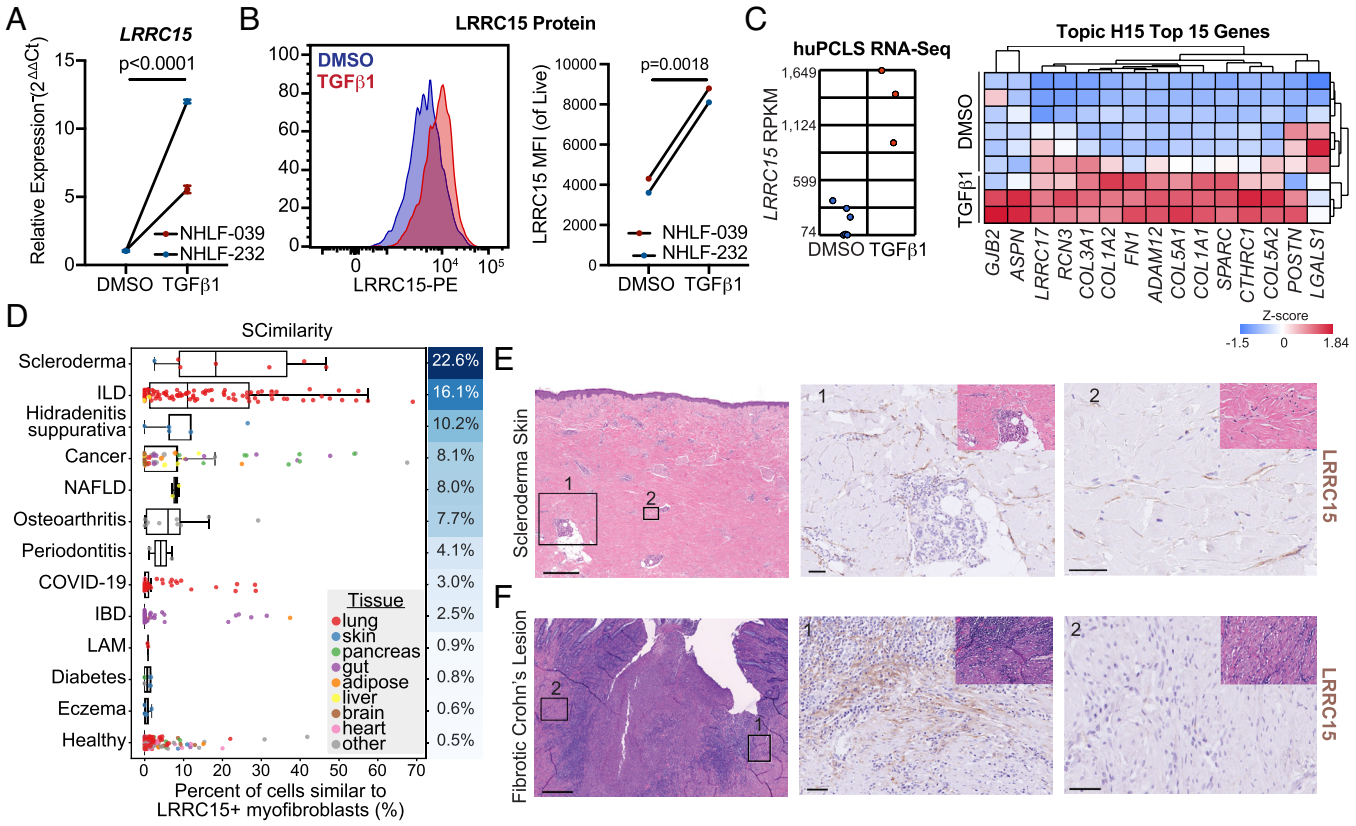

Fig. 4. TGFb promotes LRRC15+ myofibroblast gene expression found in fibrotic diseases and malignancies across tissues. (*A*) Relative expression of *LRRC15* in DMSO and TGFb1 stimulated NHLFs. (*B*) Representative flow cytometry (*Left*) and quantified cell surface protein expression (*Right*) for DMSO and TGFb1 stimulated NHLF. (*C*) Expression of *LRRC15* and top 15 genes of H15 in DMSO and TGFb1 stimulated human precision cut lung slices (huPCLS). (*D*) Percentage of cells within each sample (x-axis) similar to *LRRC15*+ myofibroblasts across diseases (y-axis, *Left*) and tissues (point color) sorted by mean percentage of similar cells across samples (color scale, *Right*). Restricted to samples with ≥50 fibroblast lineage cells and diseases with ≥0 similar cells. (*E* and *F*) Immunohistochemistry staining for LRRC15 with *Inset* H&E in (*E*) scleroderma skin and (*F*) Crohn's Disease intestine.

adenocarcinoma (Fig. 4*D*). These results suggest that convergent transcriptional programs associated with LRRC15+ myofibroblasts are activated in malignant and nonmalignant disease settings and that the findings herein may be applicable beyond IPF.

Last, to determine if the SCimilarity results are indicative of LRRC15 protein expression in nonmalignant disease tissue niches outside of the lung, we evaluated LRRC15 expression by IHC in samples derived from patients with Crohn's Disease (CD) and Scleroderma (SSc), two indications predicted to harbor cells similar to LRRC15-expressing myofibroblasts in IPF. In SSc, a chronic autoimmune disorder characterized by progressive fibrosis of the skin and internal organs, LRRC15 protein expression was detected in spindle shaped cells embedded within thickened bundles of dermal collagen (Fig. 4*E*). In donor-derived samples of Crohn's ileitis, a form of inflammatory bowel disease in which fibrostenosis presents a severe complication, LRRC15 protein localized to areas of active fibrosis adjacent to mucosal ulceration and mural fistulae (Fig. 4*F*). Notably, LRRC15 protein is not detected in areas of mature fibrosis, suggesting that its expression is restricted to foci of active fibroplasia, similar to the restricted spatial distribution in human IPF lung tissue.

Discussion

Here, we identify LRRC15 as a cellular marker of FF-associated myofibroblasts in IPF. To elucidate the heterogeneity, gene expression networks, and regulatory determinants of profibrotic lung myofibroblasts, we extensively profiled fibroblasts, including LRRC15-expressing myofibroblasts, in a human IPF fibroblast scRNA-seq atlas and a mouse model of injury-induced lung fibrosis. We identified two novel conserved myofibroblast expression programs that are enriched for genes associated with either wound healing and inflammation or extracellular matrix production and modification. We predict that inflammatory IL-1b and TNFa, growth factor signaling, and profibrotic TGFb and Wnt signaling pathways are upstream regulators of these two phenotypes using in silico, in vitro, and in vivo approaches. We mechanistically demonstrate that fibroblast TGFb signaling is a required regulator of the cell-state transition from a proliferative and inflammation associated to an ECM producing myofibroblast. Last, using a recently developed single-cell foundation model, we queried a pan-tissue and -disease cellular space using IPF LRRC15+ myofibroblasts and found similar cell states across human fibrotic and malignant conditions, suggesting the potential for broad applicability of the findings herein.

Previous studies have demonstrated the central role TGFb plays in wound healing, fibrosis, and the differentiation of myofibroblasts in a variety of contexts including lung fibrosis (37, 38). Using a lineage-tracing conditional knockout model of *Tgfbr2*, we demonstrated here the requirement for fibroblast-specific TGFb signaling for LRRC15-myofibroblast differentiation. Novel lineage-specific genetic mouse model tools based on *Cdh11* validated our hypothesis, revealing the absolute mechanistic requirement for TGFbR2 signaling in alveolar and activated fibroblasts following fibrosis-inducing injury. Critically, these studies also identified a role for TGFb signaling in directly

suppressing wound-healing and immune-interacting gene programs, and conditional knockout of the *Tgbfr2* in CDH11-expressing fibroblasts resulted in alterations in the immune composition of injured lung tissue. These results raise critical questions regarding the therapeutic effects of inhibiting profibrotic signaling pathways in human disease, where the relative contribution of inflammation and ECM remodeling to pathophysiology is incompletely understood. A marker such as LRRC15, which is tightly associated with a functional gene program, may provide a framework to study this question using genetic models in future preclinical studies.

TGFb signaling is a critical pleiotropic regulator of inflammation (39). In immune cells, TGFb signaling promotes regulatory and tissue-regenerative phenotypes often desirable to reduce pathological inflammation in chronic disease. In stromal cells, TGFb signaling promotes wound-healing gene programs that, when chronically activated, lead to pathological fibrosis and loss of organ function (40). This trade-off between pathological inflammation and fibrosis is a subject that warrants further investigation and may be a primary barrier to the success of TGFb targeted therapeutics, which have had little success despite promising preclinical data.

The identification of unified cellular markers that associate with pathogenic cellular programs is highly desired. Such identifiers enable research through the application of genetically engineered mouse models and have potential therapeutic utility. Several markers have been identified for emergent myofibroblasts in perturbed tissue conditions, including LRRC15, CTHRC1, aSMA, POSTN, and FAP and several recent reports have identified two or more myofibroblast subtypes across perturbed tissue contexts in the lung, skin, liver, kidney, and heart (6, 8, 41–43). Although these markers overlap in some contexts, considerable differences in expression patterns exist that warrant granular characterization. Myofibroblast subsets may derive from separate precursors, occupy different tissue niches or represent distinct phenotypes and biological functions along a differentiation trajectory. For example, in our mouse studies we note a previously undescribed SMC-derived population of LRRC15-positive cells, a finding that warrants further investigation in mouse models and human data. Our findings highlight that LRRC15 is a highly restricted marker of fibrosis-associated myofibroblasts found in pathological niches across human diseases. Importantly, we find that the core transcriptional program of LRRC15+ myofibroblasts can be activated in human tissue organoids in the absence of an intermediate cell state. This is in contrast to other markers, including CTHRC1, that we find upregulated either earlier in the differentiation trajectory or in emergent wound responsive fibroblasts that differentiate through an inflammatory intermediate (9). This observation may inform strategies that aim to cure fibrosis by focusing on limiting inflammation. Moreover, this insight may be essential in understanding whether fibrosis is reversible.

Limitations of the Study

There are several important limitations to this study. The scarce availability of IPF patient tissue prior to end stage disease renders studying myofibroblast subcategorization and development with high confidence quite challenging. To address this, we employed a mouse model of human disease with serial sampling and in silico disease progression modeling approaches. Understanding the mechanisms that underlie early, progressive and treatment-resistant disease will be critical to developing novel therapies for IPF. We also make use of several computational analyses, including Milo-DA/DE and Slingshot trajectory, to make predictions regarding the putative progressive nature of the differentiation states we identify. It is important to note that these predictions require experimental validation using appropriate lineage tracing tools, such as an *Lrrc15*.CreERT2 mouse model. In addition, while the SCimilarity tool spans a large number of public sc/snRNA-seq datasets across the entire human body, it is not exhaustive of all diseases and can only reveal insights regarding published studies included in the corpus on the date of model training. Nevertheless, these results demonstrate that a similar cellular program to LRRC15+ myofibroblasts is present in diverse disease states and tissues, implying potential therapeutic relevance in multiple fibrotic and malignant indications.

Materials and Methods

Methods.

Mice. All mouse experiments were performed under protocols approved by the Institutional Animal Care and Use Committee of Genentech, Inc or the Animal Care Committee at The Centre for Phenogenomics, Toronto, ON.

Dpt^{IresCreERT2} mice (6) were designed, generated, and bred at Genentech. Wild-type C57/B6J (000664), Tgfbr2.flox (012603), and Rosa26.LSL.YFP (006148) mice were obtained from The Jackson Laboratory. For bleomycin studies, age-matched males (15 to 25 wk old at start of study) were used for all studies. Mice were maintained under specific pathogen-free conditions using the guidelines of the US NIH.

Mouse tissue digestion and cell isolation and flow cytometry. Mouse lung cells were isolated from total lung tissue or selected lobes as previously described (Buechler, Immunity, 2019.) Briefly, lung tissue was manually minced using scissors and incubated in enzyme-containing digestion media (RPMI + 2% FBS with 100 mg/mL Dispase (Life Tech., Cat #17105041), 100 mg/mL collagenase P (Roche, Cat #11249002001) and 50 mg/mL DNase I (Roche, Cat # 10104159001) for three incubation periods of 12 min at 37 °C. Between each incubation period, the tissue was agitated by pipetting the sample repeatedly and the cell containing supernatant removed, filtered (70 μM), placed on ice, and replaced with fresh digestion media. After digestion, samples were incubated with ACK lysis buffer for 2 min to remove red blood cells prior to downstream applications.

Bleomycin histopathology. For PITI bleomycin characterization, formalin fixed whole lung or lung lobes were processed in a Tissue Tek VIP 6 tissue processor (Sakura Finetek, USA), paraffin-embedded, 4 μm-thick sections were collected onto charged slides (Assure, Epic Scientific, USA) and stained with hematoxylin & eosin or Masson's trichrome in a Tissue Tek Prisma autostainer (Sakura Finetek, USA). Slides were scanned at 20× using an Olympus VS120 slide scanner equipped with a Hamamatsu ORCA-R2 C10600 digital camera (Evident, Japan).

All slides were randomly numbered in a blinded fashion and scored by an anatomical pathologist, using the modified Ashcroft scale (44). Briefly, the lungs were inspected microscopically with a 20× objective, following a raster-like pattern throughout the sections. Foci with predominant trachea or bronchial mucosa were excluded. Each 20× focus was graded, summarized, and then divided by the total number of fields to obtain a fibrotic index for the lung.

LRRC15 Immunohistochemistry and Immunofluorescence. LRRC15 protein was detected by immunohistochemistry in formalin-fixed, paraffin embedded tissue using standard Thermo Scientific Autostainer protocols. Briefly, 4-μm sections were subject to antigen retrieval (Target retrieval) and incubated with a primary anti-Lrrc15 antibody (Abcam clone EPR8188; Cat # ab150376), followed by a secondary antibody (biotinylated anti-rabbit IgG) and a tertiary reagent (ABC-peroxidase elite), then visualized with DAB detection. Images were acquired from scanned slides using the Nanozoomer S360 automated digital slide scanning platform equipped with 20× (NA 0.75) Nikon lens (Hamamatsu Photonics).

LRRC15 protein was detected by immunofluorescence in agarose inflated (1% in PBS) lung tissue. Tissue was fixed in 4% PFA in PBS at room temperature for 3 to 4 h. Tissue was then incubated overnight sequentially in 20% and 30% sucrose in PBS at 4C. Tissue was embedded in optimal cutting temperature medium and frozen for cryosectioning. Slides were stained using the Epridia Sequenza

Capillary staining system. Briefly, slides were incubated with primary anti-LRRC15 antibody (M25, 10 ng/mL) overnight at 4C and secondary antibody (Goat anti-human Alexa 488 Cat # A11013, 1:1,000 dilution) for 1 h at RT. Slides were imaged on an Olympus BX61 Epifluorescence Microscope.

In vitro human fibroblast and PCLS cell culture and sample preparation. NHLFs from multiple donors (Lonza) or human precision cut lung slices (huPCLS; Institute for In Vitro Sciences) were cultured in 10% FBS DMEM. For cytokine and small molecule treatment, cells were treated with TGFb-1 (5 ng/mL, R&D Cat # 7666-MB-005/CF) for 72 h prior to flow cytometry staining or RNA isolation using the Qiagen RNeasy Mini Kit (Cat # 74104).

In vitro huPCLS RNA-seq analysis. Normalized counts per million (CPM) values were computed as a measure of gene expression using the method in the DESeq2 v1.40.2 R package (45, 46). Pairwise differential expression analysis between treatment conditions was performed using the limma+voom method provided by the limma v3.56.1 R package (47).

Human lung sample collection. Research participants donating human lung samples presented in this study were recruited from patients undergoing surgical lung resection at University of California San Francisco (UCSF)-affiliated medical centers including San Francisco Veterans Affairs (VA) Health Care System (SFVAHCS) and Zuckerberg San Francisco General Hospital Medical Center. The procurement of human lung tissue specimens used in this study was conducted in accordance with the Declaration of Helsinki. The UCSF Institutional Review Board (IRB) and the SFVAHCS Committee on Research and Development approved the study protocols. Written IRB-approved informed consent and Health Insurance Portability and Accountability Act were obtained from all study participants. Study participants received a compensation of up to $100 if they completed all study procedures, which included completing respiratory questionnaires, undergoing pulmonary function testing, donating a blood sample, and donating part of their surgically resected lung tissue, which was performed as part of their clinical care. The compensation amount was deemed appropriate by IRB to avoid undue influence on participation.

Human lung fibroblast scRNA-seq atlas data collection and data processing. FASTQ files were downloaded from the Sequence Read Archive (SRA) for GSE132771 (10), GSE135893 (14), GSE136831 (15), GSE173896 (16), GSE196638 (48), GSE227136 (18), and GSE159354 (19). For both scRNA-seq generated as part of this study and public scRNA-seq data downloaded from SRA, sequencing reads were aligned to the GENCODE 27 Basic gene model (49) on the human genome assembly GRCh38 using Cell Ranger v7.1.0 (10× Genomics, Pleasanton, CA). Ambient RNA and random barcode swaps were corrected using CellBender v0.3.0 (49, 50). Further analysis was restricted only to samples with a diagnosis of either normal or IPF.

Data, Materials, and Software Availability. Human scRNA-seq source data that was reanalyzed in this manuscript are available from the Sequence Read Archive (SRA) under Gene Expression Omnibus (GEO) accessions GSE132771, GSE135893, GSE136831, GSE173896, GSE196638, GSE227136, and GSE159354 (10, 14–16, 18–20, 48). Human and mouse scRNA-seq and bulk RNA-seq data generated as part of this study are available from GEO under accessions GSE324380, GSE296912, GSE296913, GSE296916, GSE296918, and GSE296920 (51–56).

ACKNOWLEDGMENTS. We thank contributors from the Research Pathology Labs at Genentech, Inc. and The Centre for Phenogenomics Pathology Core. We thank H. Jackson for CyTOF isotope conjugation. This work was supported by funding from Canadian Institute of Health Research Project Grant 471606 (M.B.B.) and Genentech, Inc.

Author affiliations: [a]Genentech Research and Early Development, Genentech, Inc., South San Francisco, CA 94080; [b]Roche Informatics, F. Hoffmann-La Roche Ltd., Mississauga, ON L5N 5M8, Canada; [c]Department of Immunology, University of Toronto, Toronto, ON M5S 3G9, Canada; [d]Hospital for Sick Children Research Institute, Toronto, ON M5G 1E8, Canada; [e]St. Joseph's Health Centre, Unity Health Toronto, Toronto, ON M6R 1B5, Canada; [f]Department of Medicine, University of California San Francisco, San Francisco, CA 94143; and [g]Schroeder Arthritis Institute, University Health Network, Toronto, ON M5G 2C4, Canada

1. M. Zeisberg, R. Kalluri, Cellular mechanisms of tissue fibrosis. 1. common and organ-specific mechanisms associated with tissue fibrosis. *Am. J. Physiol. Cell Physiol.* **304**, C216–C225 (2013).
2. M. L. Smith, The histologic diagnosis of usual interstitial pneumonia of idiopathic pulmonary fibrosis. Where we are and where we need to go. *Mod. Pathol.* **35**, 8–14 (2022).
3. D. S. Glass *et al.*, Idiopathic pulmonary fibrosis: Current and future treatment. *Clin. Respir. J.* **16**, 84–96 (2022).
4. L. Richeldi *et al.*, Nerandomilast in patients with idiopathic pulmonary fibrosis. *N. Engl. J. Med.* **392**, 2193–2202 (2025).
5. R. T. Kendall, C. A. Feghali-Bostwick, Fibroblasts in fibrosis: Novel roles and mediators. *Front. Pharmacol.* **5**, 123 (2014).
6. M. B. Buechler *et al.*, Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
7. I. Korsunsky *et al.*, Cross-tissue, single-cell stromal atlas identifies shared pathological fibroblast phenotypes in four chronic inflammatory diseases. *Med* **3**, 481–518.e14 (2022).
8. Y. Gao *et al.*, Cross-tissue human fibroblast atlas reveals myofibroblast subtypes with distinct roles in immune modulation. *Cancer Cell* **42**, 1764–1783.e10 (2024).
9. T. Tsukui, P. J. Wolters, D. Sheppard, Alveolar fibroblast lineage orchestrates lung inflammation and fibrosis. *Nature* **631**, 627–634 (2024).
10. T. Tsukui *et al.*, Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat. Commun.* **11**, 1920 (2020).
11. Z. Mukhatayev, A. Adilbayeva, J. Kunz, CTHRC1: An emerging hallmark of pathogenic fibroblasts in lung fibrosis. *Cells* **13**, 946 (2024).
12. C. X. Dominguez *et al.*, Single-cell RNA sequencing reveals stromal evolution into LRRC15 myofibroblasts as a determinant of patient response to cancer immunotherapy. *Cancer Discov.* **10**, 232–253 (2020).
13. J. W. Purcell *et al.*, LRRC15 is a novel mesenchymal protein and stromal target for antibody-drug conjugates. *Cancer Res.* **78**, 4059–4072 (2018).
14. A. C. Habermann *et al.*, Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).
15. T. S. Adams *et al.*, Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, eaba1983 (2020).
16. N. Watanabe *et al.*, Anomalous epithelial variations and ectopic inflammatory response in chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* **67**, 708–719 (2022).
17. K. Natsuga *et al.*, Cas9-guided haplotyping of three truncation variants in autosomal recessive disease. *Hum. Mutat.* **43**, 877–881 (2022).
18. H. M. Natri *et al.*, Cell-type-specific and disease-associated expression quantitative trait loci in the human lung. *Nat. Genet.* **56**, 595–604 (2024).
19. D. J. DePianto *et al.*, Molecular mapping of interstitial lung disease reveals a phenotypically distinct senescent basal epithelial cell population. *JCI Insight* **6**, e143626 (2021).
20. M. Eyres *et al.*, Spatially resolved deconvolution of the fibrotic niche in lung fibrosis. *Cell Rep.* **40**, 111230 (2022).
21. A. Moeller, K. Ask, D. Warburton, J. Gauldie, M. Kolb, The bleomycin animal model: A useful tool to investigate treatment options for idiopathic pulmonary fibrosis? *Int. J. Biochem. Cell Biol.* **40**, 362–382 (2008).
22. C. H. Mayr *et al.*, Sfrp1 inhibits lung fibroblast invasion during transition to injury-induced myofibroblasts. *Eur. Respir. J.* **63**, 2301326 (2024).
23. K. Street *et al.*, Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 1–16 (2018).
24. D. Correa-Gallegos *et al.*, CD201+ fascia progenitors choreograph injury repair. *Nature* **623**, 792–802 (2023).
25. W. Cao *et al.*, Cadherin-11 deficiency attenuates Ang-II-induced atrial fibrosis and susceptibility to atrial fibrillation. *J. Inflamm. Res.* **14**, 2897–2911 (2021).
26. D. J. Schneider *et al.*, Cadherin-11 contributes to pulmonary fibrosis: Potential role in TGF-β production and epithelial to mesenchymal transition. *FASEB J.* **26**, 503–512 (2012).
27. G. Zhang, X. Wang, Q. Zhang, Cdh11: Roles in different diseases and potential value in disease diagnosis and treatment. *Biochem. Biophys. Rep.* **36**, 101576 (2023).
28. B. Hinz, P. Pittet, J. Smith-Clerc, C. Chaponnier, J.-J. Meister, Myofibroblast development is characterized by specific cell-cell adherens junctions. *Mol. Biol. Cell.* **15**, 4310–4320 (2004).
29. S. K. Chang *et al.*, Cadherin-11 regulates fibroblast inflammation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8402–8407 (2011).
30. M. Pedroza, S. To, J. Smith, S. K. Agarwal, Cadherin-11 contributes to liver fibrosis induced by carbon tetrachloride. *PLoS One* **14**, e0218971 (2019).
31. M. Lodyga *et al.*, Cadherin-11-mediated adhesion of macrophages to myofibroblasts establishes a profibrotic niche of active TGF-β. *Sci. Signal.* **12**, eaao3469 (2019).
32. H. Higashiyama *et al.*, Inhibition of activin receptor-like kinase 5 attenuates bleomycin-induced pulmonary fibrosis. *Exp. Mol. Pathol.* **83**, 39–46 (2007).
33. T. Sun *et al.*, TGFβ2 and TGFβ3 isoforms drive fibrotic disease pathogenesis. *Sci. Transl. Med.* **13**, eabe0407 (2021).
34. J. Zhao *et al.*, Smad3 deficiency attenuates bleomycin-induced pulmonary fibrosis in mice. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **282**, L585–L593 (2002).
35. A. T. Krishnamurty *et al.*, LRRC15+ myofibroblasts dictate the stromal setpoint to suppress tumour immunity. *Nature* **611**, 148–154 (2022).
36. G. Heimberg *et al.*, A cell atlas foundation model for scalable search of similar human cells. *Nature* **638**, 1085–1094 (2024).
37. N. Frangogiannis, Transforming growth factor-β in tissue fibrosis. *J. Exp. Med.* **217**, e20190103 (2020).
38. S. O'Kane, M. W. Ferguson, Transforming growth factor beta s and wound healing. *Int. J. Biochem. Cell Biol.* **29**, 63–78 (1997).
39. E. Battle, J. Massague, Transforming Growth Factor-β signaling in immunity and cancer. *Immunity* **50**, 924–940 (2019).
40. A. Biernacka, M. Dobaczewski, N. G. Frangogiannis, TGF-β signaling in fibrosis. *Growth Factors* **29**, 196–202 (2011).
41. C. F. Guerrero-Juarez *et al.*, Single-cell analysis reveals fibroblast heterogeneity and myeloid-derived adipocyte progenitors in murine skin wounds. *Nat. Commun.* **10**, 650 (2019).
42. O. Krenkel, J. Hundertmark, T. P. Ritz, R. Weiskirchen, F. Tacke, Single cell RNA sequencing identifies subsets of hepatic stellate cells and myofibroblasts in liver fibrosis. *Cells* **8**, 503 (2019).
43. N. Farbehi *et al.*, Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. *Elife* **8**, e43882 (2019).

44. R.-H. Hübner *et al.*, Standardized quantification of pulmonary fibrosis in histological samples. *Biotechniques* **44**, 514–517 (2008).
45. P. Badia-I-Mompel *et al.*, decoupleR: Ensemble of computational methods to infer biological activities from omics data. *Bioinform. Adv.* **2**, vbac016 (2022).
46. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
47. M. E. Ritchie *et al.*, Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
48. C. Wang *et al.*, Dysregulated lung stroma drives emphysema exacerbation by potentiating resident lymphocytes to suppress an epithelial stem cell reservoir. *Immunity* **56**, 576–591.e10 (2023).
49. A. Frankish *et al.*, GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
50. S. J. Fleming *et al.*, Unsupervised removal of systematic background noise from droplet-based single-cell experiments using Cell Bender. *Nat. Methods* **20**, 1323–1335 (2023).
51. C. D. Davidson *et al.*, scRNA-seq of fibroblasts from inducible Cdh11 reporter mice. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE324380. Deposited 9 March 2026.
52. G. Teng *et al.*, scRNA-seq of normal and COPD lungs. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE296912. Deposited 12 May 2025.
53. C. Cox *et al.*, Transcriptional profiling of mouse fibroblasts stimulated with cytokines in vitro. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE296913. Deposited 12 May 2025.
54. J. A. Shyer *et al.*, scRNA-seq of PDPN+ fibroblasts from various timepoints from a mouse model of fibrotic lung disease. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE296916. Deposited 12 May 2025.
55. J. A. Shyer *et al.*, scRNA-seq assessment of the impact of loss of Tgfbr2 in DPT-expressing fibroblasts in a model of fibrotic lung disease. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE296918. Deposited 12 May 2025.
56. J. A. Shyer *et al.*, scRNA-seq of lung cells sorted from the acute 3X IT bleomycin induced fibrosis mouse model. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE296920. Deposited 12 May 2025.