ATAC-ing single nucleus in idiopathic pulmonary fibrosis: TWIST1 strives back for myofibroblasts

Olivier Burgy[1,2] and Arnaud A. Mailleux[3]

[1]INSERM U1231, Faculty of Medicine and Pharmacy, University of Bourgogne-Franche Comté, Dijon, France. [2]Constitutive Reference Center for Rare Pulmonary Diseases – OrphaLung, Dijon-Bourgogne University Hospital, Dijon, France. [3]Université Paris Cité, Inserm, Physiopathologie et épidémiologie des maladies respiratoires, F-75018 Paris, France.

Corresponding author: Arnaud A. Mailleux (arnaud.mailleux@inserm.fr)

Shareable abstract (@ERSpublications)
Integrated multi-omic single cell analyses in IPF lungs highlights TWIST1 as a critical transcriptional regulator for myofibroblasts https://bit.ly/3qBjh0S

Cite this article as: Burgy O, Mailleux AA. ATAC-ing single nucleus in idiopathic pulmonary fibrosis: TWIST1 strives back for myofibroblasts. *Eur Respir J* 2023; 62: 2300881 [DOI: 10.1183/13993003.00881-2023].

Received: 25 May 2023
Accepted: 5 June 2023

Ageing is often associated with chronic remodelling, which is a key feature of many human diseases, including chronic respiratory diseases such as lung fibrosis. Idiopathic pulmonary fibrosis (IPF) is the most common form of pulmonary fibrosis and is defined as a specific form of chronic, progressive fibrosing interstitial pneumonia of unknown cause. IPF patients have an overall median survival of 3 to 5 years, although current antifibrotics (pirfenidone and nintedanib) slow down lung function decline and may improve survival [1, 2]. The current paradigm suggests that IPF is a lung disorder of epithelial origin in which aberrant epithelial cell activation and ineffective repair lead to aberrant mesenchymal responses and accumulation, with secondary deposition of extracellular matrix (ECM) in the alveolar interstitium: one of the salient marks of fibrosis [1, 3, 4]. Fibroblasts are the main mesenchymal components responsible for tissue remodelling during normal and pathological lung tissue repair. In IPF, chronic injuries may lead to unbalanced cell–cell communication among the lung cells that would promote the persistence of activated "pathological" fibroblasts or myofibroblasts, and hence excessive and deleterious accumulation of ECM [1, 3]. Myofibroblasts have been believed for a long time now to constitute a heterogenous population with many possible cellular origins and several effector properties. Numerous studies utilising single-cell RNA sequencing (scRNA-seq) have demonstrated the existence of various mesenchymal cell lineages within the lung. These studies have also brought attention to the emergence of novel abnormal mesenchymal cell types in both IPF lung samples and preclinical mouse models of lung fibrosis [2, 5–7].

Cell differentiation and activation are characterised by significant changes in gene expression patterns, which are primarily driven by chromatin remodelling. One of the key processes involved in chromatin remodelling is the alteration of chromatin structures from tightly condensed state to more relaxed, transcriptionally permissive conformation. This enables the binding of transcription factors (TFs) and other DNA-binding proteins to the DNA, allowing for the specific regulation of gene expression [8, 9]. Through aberrant changes in chromatin landscape, mesenchymal cells would differentiate and acquire dysregulated functions in IPF.

In the current issue of the *European Respiratory Journal*, Valenzi *et al.* [10] present exciting new insights on the activity of critical TFs associated with "pathological" myofibroblasts in IPF, using multi-omic single-cell analyses. To shed light on the TFs governing the activation and differentiation of myofibroblasts, they leveraged cutting edge genomic methodology at the single cell resolution (figure 1). When it comes to epigenetics, accessible chromatin often translates into expressed genome areas as it allows TFs to access DNA. In 2013, the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) was introduced and allowed unbiased screening of accessible chromatin at the single cell resolution [11, 12]. The nuclear genome composed of histone-bound and -unbound DNA is exposed to hyperactive Tn5 transposase. As illustrated in figure 1c, the enzyme breaks and releases DNA sequences

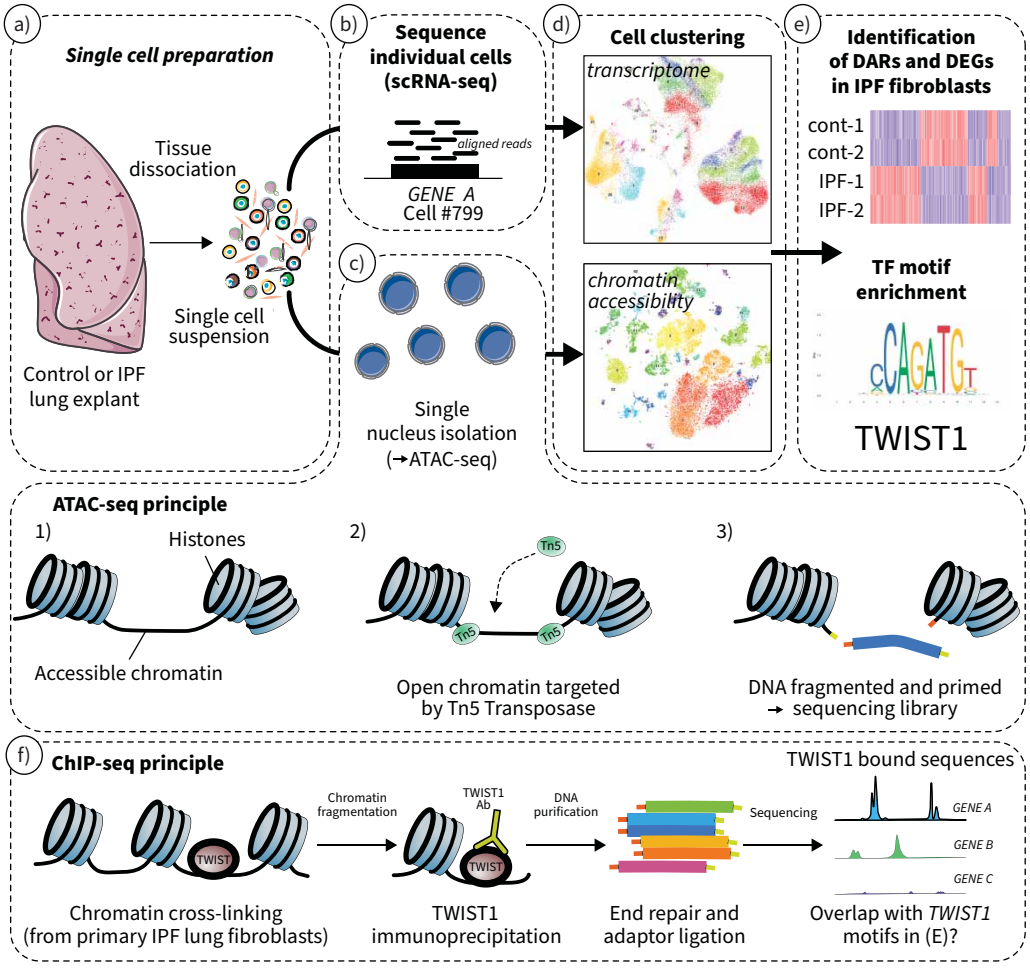


FIGURE 1 Schematic summary of the workflow leading to TWIST1 identification using single cell RNA sequencing (scRNA-seq) and single nucleus Assay for Transposase-Accessible Chromatin using sequencing (snATAC-seq) multi-omics in idiopathic pulmonary fibrosis (IPF) fibroblast clusters. a) Single cell suspensions from control or IPF lung explants were divided to generate b) scRNA-seq or c) snATAC-seq datasets after nucleus isolation for the latter. The ATAC-seq method involves exposing accessible chromatin regions (1) to the highly active Tn5 transposase (2). Tn5 simultaneously fragments DNA, preferentially inserts into open chromatin sites, and adds sequencing primers (3). The fragmented and primed DNA identifies the open chromatin after subsequent sequencing library generation and analysis. d) The datasets generated in scRNA-seq and snATAC-seq were used to identify and cluster lung cell types. The clusters can be visualised using a UMAP diagram, as shown. e) Integration of both scRNA-seq (transcriptome) and snATAC-seq (chromatin accessibility) will identify differentially accessible chromatin regions (DARs) and differentially expressed genes (DEGs) enriched in IPF fibroblast lung clusters. Subsequent analysis showed an enriched presence of transcription factor (TF)-binding motifs for TWIST1 within these DEG-accessible chromatin regions in IPF fibroblast clusters. f) To validate the TWIST1 binding motifs, chromatin immunoprecipitation-sequencing (ChIP-seq) was performed using a TWIST1 antibody (Ab) on bulk cross-linked chromatin prepared from cultured primary IPF lung fibroblasts. After genome mapping, the coverage of TWIST1 Ab bound sequence was then compared with the active TWIST1 motifs identified in e. Panels a and c were drawn using pictures from Servier Medical Art, which is licensed under a Creative Commons Attribution 3.0 Unported License. The TWIST1 sequence logo in panel e was obtained from the JASPAR database (jaspar.genereg.net).

which are not linked to histone, favouring the eventual binding of transcription regulators. High-throughput sequencing of those DNA sequences allows the identification of differentially accessible chromatin regions (DARs) within the samples. The accessible DNA sequences identified are then used as input to a motif database to uncover which TFs were likely bound into the isolated sequences. Valenzi *et al.* [10] applied this methodology to single nuclei (sn) isolated from the lung of patients with IPF, as well as donors, to

snapshot the landscape of accessible chromatin in the major lung cell types. As a step further, each patient sample underwent, in parallel, single-cell transcriptomics in order to get both ATAC- and RNA-seq data for each patient. This allows the cell annotation of the snATAC-seq and subclustering of the mesenchymal populations, on which the downstream investigation of the study was focused. Comparison of IPF myofibroblasts *versus* non-myogenic fibroblasts or fibroblasts from donors shows that accessible chromatin areas were located to genes involved in fibrosis-relevant pathways. Deep analysis of accessible chromatin within the mesenchymal populations also highlighted TWIST1 as a key transcriptional regulator in the pathological myofibroblast phenotype observed in IPF. TWIST1 belongs to the basic helix–loop–helix superfamily of TFs, which play a crucial role in controlling gene expression and are essential for lineage specification during development [13]. TWIST1 binds DNA directly, through recognition of variants of a DNA motif called E-box sequence [13]. This motif was highly enriched in fibroblasts compared to other cells in the study by Valenzi *et al.* [10]. A chromatin immunoprecipitation-sequencing assay was also conducted in cultured primary IPF fibroblasts to identify the DARs that directly interacted with TWIST1 (figure 1f). The authors showed a good coverage of DARs by TWIST1 bound sequences, despite the biases potentially associated with *in vitro* culture. In addition, TWIST1 was among the few TFs upregulated and identified as differentially expressed genes between donor and IPF myofibroblast populations. To further validate their findings and explore the functional role of this TF in lung fibrosis, the authors performed gain-of-function studies using genetically modified mice overexpressing *Twist1* in *Col1a2*-expressing cells. Increased *Twist1* expression in collagen-producing cells was associated with an augmentation in collagen synthesis. This effect was observed in both *in vitro* experiments using primary lung fibroblasts and *in vivo* employing the bleomycin model of lung fibrosis.

While the study from Valenzi *et al.* [10] undoubtedly adds significant insight on the epigenetics of mesenchymal cells during IPF, some questions remain with regard to the biology of TWIST1 in IPF. The role of TWIST1 in lung fibrosis has been previously reported by different groups in the past decades. While TWIST1 is primarily associated with aberrant myofibroblast activation in lung fibrosis [14–17], this TF is also expressed by other lung cell types, such as remodelled epithelium [16] and endothelial cells [18], as well as macrophages [16] in pulmonary fibrosis. Of note, TWIST1 is well known for its involvement in the epithelial-to-mesenchymal transition during cancer metastasis [13]. Their integrated scRNA-seq and snATAC-seq methodologies highlighted an enrichment in TWIST1 motifs in IPF endothelial but not epithelial cell nuclei. However, the authors acknowledged that the constrained sample size of nuclei and the heterogeneous nature of epithelial lung populations may have attenuated the power of their method. In addition, some of the samples were depleted for immune and epithelial cells prior to snATAC-seq in this study. Further investigations are needed to explore whether TWIST1 is involved in other non-mesenchymal cells during IPF.

In summary, the “tour de force” of this study resides in the integration of snATAC-seq and scRNA-seq datasets to identify differentially accessible chromatin regions and enriched TF motifs within lung cell populations, and to highlight the involvement of TWIST1 in lung fibrosis in the process and by using *in vitro* and *in vivo* preclinical models of lung fibrosis. Previous studies targeting TFs in lung fibrosis, including GLI [19], FOXM1 [20], FOXF1 [21], FOXO3 [22], RUNX2 [23], PU.1 [24] and TBX4 [24, 25], suggested that such therapeutic approaches could be beneficial in this chronic lung disease. However, with the exception of TBX4, the expression of these TFs is not restricted to mesenchymal lineages. Therefore, targeting these TFs may have implications for both fibrosis development and epithelial regeneration/repair. The potential development of small molecule therapeutics that inhibit DNA-protein binding and TF complexes would help to target TWIST1 dimerisation pairs or its association with binding partners. Indeed, the diversity of TWIST1 dimer combinations and their functional specificity in lung cell types, which was not explored by Valenzi *et al.* [10], remains elusive. Therefore, the identification of TWIST1 partners specifically in pathological myofibroblasts (as opposed to the epithelium or the endothelium) would facilitate potential targeting of this aberrant pathological mesenchymal population.

Finally, it remains unclear which nuclear factors (including TWIST1 itself?) in addition to known epigenetic processes, such as DNA methylation [26] and histone modifications [27], contribute to the promotion of those newly accessible chromatin regions that are associated with the aberrant phenotype of activated myofibroblasts observed in IPF. Furthermore, recent efforts have been made to map DNA methylation [28] and histone marks [29] at single-cell resolution. These developments are noteworthy as they could shape forthcoming multi-omics investigations in IPF. Over the past years, omics analysis applied to IPF or experimental models has significantly deepened our knowledge of pulmonary fibrosis. As time passes, innovative multi-dimensional methodologies are developing rapidly. While currently exploring both transcriptomic and chromatin accessibility implies two separate workflows and specific analysis for inferred linkage, novel methods now enable both ATAC- and RNA-seq simultaneously on the

same cell. In the future, technology breakthroughs allowing simultaneous analysis of epigenetics (DNA methylation, histone modification), at the single cell resolution, will be important to further deepen our understanding of IPF. Uncovering the epigenetic mechanisms involved in chromatin remodelling, along with gene expression regulation in activated myofibroblasts, is crucial for the development of effective therapies that target the underlying mechanisms of IPF and hopefully promote proper repair of the lung.

Conflict of interest: The authors have no potential conflicts of interest to disclose.

References

1 Froidure A, Marchal-Duval E, Homps-Legrand M, *et al.* Chaotic activation of developmental signalling pathways drives idiopathic pulmonary fibrosis. *Eur Respir Rev* 2020; 29: 190140.
2 Wollin L, Distler JHW, Redente EF, *et al.* Potential of nintedanib in treatment of progressive fibrosing interstitial lung diseases. *Eur Respir J* 2019; 54: 1900161.
3 Selman M, Pardo A. When things go wrong: exploring possible mechanisms driving the progressive fibrosis phenotype in interstitial lung diseases. *Eur Respir J* 2021; 58: 2004507.
4 Burgy O, Bellaye PS, Beltramo G, *et al.* Pathogenesis of fibrosis in interstitial lung disease. *Curr Opin Pulm Med* 2020; 26: 429–435.
5 Adams TS, Schupp JC, Poli S, *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* 2020; 6: eaba1983.
6 Habermann AC, Gutierrez AJ, Bui LT, *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* 2020; 6: eaba1972.
7 Reyfman PA, Walter JM, Joshi N, *et al.* Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med* 2019; 199: 1517–1536.
8 Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell* 2007; 128: 707–719.
9 Zheng H, Xie W. The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol* 2019; 20: 535–550.
10 Valenzi E, Bahudhanapati H, Tan J, *et al.* Single-nucleus chromatin accessibility identifies a critical role for TWIST1 in idiopathic pulmonary fibrosis myofibroblast activity. *Eur Respir J* 2023; 62: 2200474.
11 Buenrostro JD, Giresi PG, Zaba LC, *et al.* Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 2013; 10: 1213–1218.
12 Buenrostro JD, Wu B, Litzenburger UM, *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 2015; 523: 486–490.
13 Ning X, Zhang K, Wu Q, *et al.* Emerging role of Twist1 in fibrotic diseases. *J Cell Mol Med* 2018; 22: 1383–1391.
14 Chen Y, Zhao X, Sun J, *et al.* YAP1/Twist promotes fibroblast activation and lung fibrosis that conferred by miR-15a loss in IPF. *Cell Death Differ* 2019; 26: 1832–1844.
15 Hanmandlu A, Zhu L, Mertens TCJ, *et al.* Transcriptomic and epigenetic profiling of fibroblasts in idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2022; 66: 53–63.
16 Bridges RS, Kass D, Loh K, *et al.* Gene expression profiling of pulmonary fibrosis identifies Twist1 as an antiapoptotic molecular 'rectifier' of growth factor signaling. *Am J Pathol* 2009; 175: 2351–2361.
17 Tan J, Tedrow JR, Nouraie M, *et al.* Loss of Twist1 in the mesenchymal compartment promotes increased fibrosis in experimental lung injury by enhanced expression of CXCL12. *J Immunol* 2017; 198: 2269–2285.
18 Hendee K, Hunyenyiwa T, Matus K, *et al.* Twist1 signaling in age-dependent decline in angiogenesis and lung regeneration. *Aging (Albany NY)* 2021; 13: 7781–7799.
19 Moshai EF, Wemeau-Stervinou L, Cigna N, *et al.* Targeting the hedgehog-glioma-associated oncogene homolog pathway inhibits bleomycin-induced lung fibrosis in mice. *Am J Respir Cell Mol Biol* 2014; 51: 11–25.
20 Penke LR, Speth JM, Dommeti VL, *et al.* FOXM1 is a critical driver of lung fibroblast activation and fibrogenesis. *J Clin Invest* 2018; 128: 2389–2405.
21 Black M, Milewski D, Le T, *et al.* FOXF1 inhibits pulmonary fibrosis by preventing CDH2-CDH11 cadherin switch in myofibroblasts. *Cell Rep* 2018; 23: 442–458.
22 Al-Tamari HM, Dabral S, Schmall A, *et al.* FoxO3 an important player in fibrogenesis and therapeutic target for idiopathic pulmonary fibrosis. *EMBO Mol Med* 2018; 10: 276–293.
23 Mummler C, Burgy O, Hermann S, *et al.* Cell-specific expression of runt-related transcription factor 2 contributes to pulmonary fibrosis. *FASEB J* 2018: 32: 703–716.
24 Wohlfahrt T, Rauber S, Uebe S, *et al.* PU.1 controls fibroblast polarization and tissue fibrosis. *Nature* 2019; 566: 344–349.
25 Xie T, Liang J, Liu N, *et al.* Transcription factor TBX4 regulates myofibroblast accumulation and lung fibrosis. *J Clin Invest* 2016; 126: 3063–3079.
26 Wang Y, Zhang L, Huang T, *et al.* The methyl-CpG-binding domain 2 facilitates pulmonary fibrosis by orchestrating fibroblast to myofibroblast differentiation. *Eur Respir J* 2022; 60: 2003697.

27 Sanders YY, Hagood JS, Liu H, *et al.* Histone deacetylase inhibition promotes fibroblast apoptosis and ameliorates pulmonary fibrosis in mice. *Eur Respir J* 2014; 43: 1448–1458.
28 Liu H, Zhou J, Tian W, *et al.* DNA methylation atlas of the mouse brain at single-cell resolution. *Nature* 2021; 598: 120–128.
29 Prompsy P, Kirchmeier P, Marsolier J, *et al.* Interactive analysis of single-cell epigenomic landscapes with ChromSCape. *Nat Commun* 2020; 11: 5702.

Single-nucleus chromatin accessibility identifies a critical role for TWIST1 in idiopathic pulmonary fibrosis myofibroblast activity

Eleanor Valenzi[1,8], Harinath Bahudhanapati[1,8], Jiangning Tan[1], Tracy Tabib[2], Daniel I. Sullivan[1], Mehdi Nouraie[1], John Sembrat[3], Li Fan[3], Kong Chen[3], Silvia Liu[4], Mauricio Rojas[5], Audrey Lafargue[6], Dean W. Felsher[7], Phuoc T. Tran[6], Daniel J. Kass[1,9] and Robert Lafyatis[2,9]

[1]Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease, Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [2]Division of Rheumatology and Clinical Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [3]Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [4]Department of Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [5]Division of Pulmonary, Critical Care and Sleep Medicine, The Ohio State University College of Medicine, Columbus, OH, USA. [6]Department of Radiation Oncology, University of Maryland School of Medicine, Baltimore, MD, USA. [7]Division of Oncology, Departments of Medicine and Pathology, Stanford University School of Medicine, Stanford, CA, USA. [8]These authors contributed equally to this work. [9]These authors contributed equally to this work.

Corresponding author: Eleanor Valenzi (valenzie@upmc.edu)

Shareable abstract (@ERSpublications)
Multiomic single-cell analyses on human IPF lungs identify a global opening of TWIST1 and other E-box motifs in IPF myofibroblasts, with *in vivo* murine models confirming a critical regulatory function for TWIST1 in IPF myofibroblast activity https://bit.ly/423aeDl

Cite this article as: Valenzi E, Bahudhanapati H, Tan J, *et al*. Single-nucleus chromatin accessibility identifies a critical role for TWIST1 in idiopathic pulmonary fibrosis myofibroblast activity. *Eur Respir J* 2023; 62: 2200474 [DOI: 10.1183/13993003.00474-2022].

This article has an editorial commentary: https://doi.org/10.1183/13993003.00881-2023

Received: 4 March 2022
Accepted: 20 April 2023

Abstract

Background In idiopathic pulmonary fibrosis (IPF), myofibroblasts are key effectors of fibrosis and architectural distortion by excessive deposition of extracellular matrix and their acquired contractile capacity. Single-cell RNA-sequencing (scRNA-seq) has precisely defined the IPF myofibroblast transcriptome, but identifying critical transcription factor activity by this approach is imprecise.

Methods We performed single-nucleus assay for transposase-accessible chromatin sequencing on explanted lungs from patients with IPF (n=3) and donor controls (n=2) and integrated this with a larger scRNA-seq dataset (10 IPF, eight controls) to identify differentially accessible chromatin regions and enriched transcription factor motifs within lung cell populations. We performed RNA-sequencing on pulmonary fibroblasts of bleomycin-injured *Twist1*-overexpressing COL1A2 Cre-ER mice to examine alterations in fibrosis-relevant pathways following *Twist1* overexpression in collagen-producing cells.

Results TWIST1, and other E-box transcription factor motifs, were significantly enriched in open chromatin of IPF myofibroblasts compared to both IPF nonmyogenic (log_2 fold change (FC) 8.909, adjusted p-value 1.82×10^{-35}) and control fibroblasts (log_2FC 8.975, adjusted p-value 3.72×10^{-28}). *TWIST1* expression was selectively upregulated in IPF myofibroblasts (log_2FC 3.136, adjusted p-value 1.41×10^{-24}), with two regions of *TWIST1* having significantly increased accessibility in IPF myofibroblasts. Overexpression of Twist1 in COL1A2-expressing fibroblasts of bleomycin-injured mice resulted in increased collagen synthesis and upregulation of genes with enriched chromatin accessibility in IPF myofibroblasts.

Conclusions Our studies utilising human multiomic single-cell analyses combined with *in vivo* murine disease models confirm a critical regulatory function for TWIST1 in IPF myofibroblast activity in the fibrotic lung. Understanding the global process of opening TWIST1 and other E-box transcription factor motifs that govern myofibroblast differentiation may identify new therapeutic interventions for fibrotic pulmonary diseases.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a devastating fibrotic lung disease resulting in architectural distortion and impaired gas exchange, ultimately progressing to respiratory failure and death in most patients. Current therapeutics have limited effect, with no approved medications convincingly improving mortality or quality of life. While the precise pathogenesis remains unknown, current paradigms suggest that repetitive microinjuries of the alveolar epithelium provoke dysregulated crosstalk with the mesenchymal compartment, leading to expansion of an activated myofibroblast population [1]. Myofibroblasts are key effectors of fibrosis by excessive deposition of extracellular matrix and by their acquired contractile capacity, resulting in distorted lung architecture [2]. In IPF, myofibroblasts are also apoptosis resistant, overcoming the normal clearance mechanisms of physiological regeneration [3]. Myofibroblasts are the primary collagen-producing cell propagating fibrosis in diverse organs, with a high disease burden ranging from the dermal and lung fibrosis of systemic sclerosis, to nephrogenic fibrosis, cirrhosis and graft-*versus*-host disease [4].

In recent years, the widespread adoption of single-cell RNA-sequencing (scRNA-seq) has produced multiple cell atlases of the human control and IPF lung, allowing precise characterisation of cell population transcriptomes [5–8]. While gene expression provides critical information on a cell's phenotype and active signalling pathways, defining upstream regulatory networks from the transcriptome alone is imprecise. Temporal control of gene expression is regulated by the cooperative interactions of *trans*-acting DNA-binding proteins with *cis*-regulatory elements within the genome, such as promoters and enhancers [9]. These sequences dictate target gene expression in a cell-type dependent manner by recruiting sequence-specific transcription factors. The advent of single-nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq) technology now allows the study of open chromatin regions in heterogeneous cell populations directly from diseased tissues, thus connecting the input regulatory signals with the output gene expression defining each population and its effector phenotype [10]. Given the central role of myofibroblasts in pulmonary fibrosis and our current lack of any drug targeting these cells, delineating their transcription factor regulation and chromatin accessibility may identify new targets for IPF and other fibroproliferative disorders.

Here we integrated snATAC-seq and scRNA-seq from human IPF and donor control explants to identify differentially accessible chromatin regions and transcription factor motifs (consensus-sequence specific binding sites) within lung cell populations. We specifically focused on IPF myofibroblasts and identified enrichment of E-box transcription factor motifs in IPF myofibroblasts compared to both IPF nonmyogenic and control fibroblasts. As the transcription factor TWIST1 is selectively expressed in IPF myofibroblasts and binds in regions of fibroblast accessible chromatin, it was particularly implicated as a positive putative regulator of IPF myofibroblast differentiation. We further investigated TWIST1 *in vitro* and in an animal model of pulmonary fibrosis. Our studies demonstrate a critical role for TWIST1 in regulating myofibroblast effector functions in IPF.

Methods

This work was approved by the institutional review board and the institutional animal care and use committee of the University of Pittsburgh.

Explanted subpleural peripheral lung tissue was digested to single-cell suspensions as described previously [11]. Single-cell suspensions were split, with a portion used for performing scRNA-seq as described previously [12], and the remaining suspension used for nuclei generation and snATAC-seq (10X Genomics). CellRanger ATAC pipeline (v1.2.0; 10X Genomics) and the R packages Signac (v1.3.0) [13], Seurat (v4.0.3), harmony (v1.0) [14] and chromVAR (v1.12.0) [15] were used for downstream analysis. Peak calling was performed by cluster using macs2 [16]. Differentially accessible regions (DARs) were calculated by logistic regression test with number of peak region fragments as a latent variable. Wilcoxon rank sum test with Bonferroni false discovery rate correction was used for differentially expressed gene (DEG) testing.

Mouse lung fibroblasts were isolated from uninjured lungs of wild-type and transgenic mice as described previously [17]. Fibroblasts were lysed for immunoblotting to measure the protein expression of TWIST1, Collagen I, α-SMA and the housekeeping protein cyclophilin A. Cells were treated with brefeldin A at 1 h prior to lysis to inhibit secretion of collagen.

Primary lung fibroblasts from Twist1-Luc, ColCre^{+}, Rosa26RTta and their littermate controls (ColCre^{+}, Rosa26RTta) were cultured (n=3) and treated with doxycycline and tamoxifen. Total RNA was purified and libraries sequenced, with CLC Genomics Workbench 11 (Qiagen) used for transcript counts, quality

control, alignment, DEGs, preliminary enrichment analysis and hierarchical clustering. Additional enrichment analyses were conducted using ingenuity pathway analysis (IPA; Qiagen). *In vitro* and *in vivo* data were analysed by robust nonparametric two-way ANOVA ("WRS2" R 4.1.2 package), with statistical analysis indicated in figure legends. The raw data have been deposited in the National Center for Biotechnology Information's Gene Expression Omnibus (GSE214085). Murine lung scRNA-seq data obtained from GSE141259. Detailed methods are presented in the supplementary material.

Results

Single-cell transcriptional and chromatin accessibility profiling in the IPF and control lung

We performed scRNA-seq and snATAC-seq on three IPF and two donor control lung tissue samples (figure 1a,b) with 8738 nuclei included for snATAC-seq analyses after filtering. Subpleural lower-lobe tissue was collected at the time of lung transplant in individuals with IPF, and from organ donors without pre-existing lung disease. Histological review of adjacent tissue showed usual interstitial pneumonia for all IPF samples (supplementary figure S1). To increase the robustness of our scRNA-seq dataset, we included 13 additional samples, for a total of 18 samples (10 IPF, eight control), with 65 179 cells included after filtering (figure 1c,d). The R packages Seurat and Signac were used for dimensional reduction, clustering, differential expression/accessibility testing and visualisation at the individual sample and aggregate dataset level [13, 18]. We annotated snATAC-seq cell types by transferring predicted labels (supplementary figure S3a) of the scRNA-seq dataset based on their transcriptomes (figure 1d), in addition to manually identifying cell types by examining gene activity matrices (a measure of chromatin accessibility within the promoter and gene). Comparison between cell-type predictions by label transfer and manual annotations indicated that all major cell types were present in both datasets and consistently identified by both methods.

We detected all major cell types within the lung with both datasets, with 260 166 accessible chromatin regions among 8738 nuclei. In snATAC-seq, cell types can be distinguished by whether DARs of the chromatin are conformationally "open" or "closed". Epithelial and macrophage clusters had the most unique DARs (figure 1e). Sequenced peak regions were annotated to the nearest gene and region of the genome (figure 1f), with the majority of peaks in distal intergenic or intronic regions [19]. The distribution of peak genomic regions was similar across cell types.

Mesenchymal profiling

To more closely examine the myofibroblasts, we subclustered scRNA-seq and snATAC-seq fibroblast, smooth muscle and pericyte populations (snATAC-seq n=844 mesenchymal nuclei; scRNA-seq n=6149 mesenchymal cells). By transcriptomes, we identified three major populations consisting of myofibroblasts, alveolar fibroblasts and adventitial fibroblasts, and a minor population referred to as *CXCL2*^{hi} fibroblasts (figure 2a). Myofibroblasts originated primarily from IPF samples, while the nonmyogenic fibroblast populations were observed in both IPF and control samples (figure 2b). The myofibroblasts were defined by upregulation of *CTHRC1*, *POSTN*, *COMP* and *COL3A1*, the alveolar fibroblasts by *SPINT2*, *FGFR4*, *GPC3* and *MACF1* (analogous to our previously described *SPINT2*^{hi} fibroblasts), and the adventitial fibroblasts by upregulation of *PI16*, *MFAP5*, *IGFBP6* (analogous to our previously described *MFAP5*^{hi} fibroblasts) (supplementary figures S4a and S5a) [11, 12, 20]. In the snATAC-seq mesenchymal subclustering, two clusters of myofibroblasts, three clusters of nonmyogenic fibroblasts and a cluster of pericytes and smooth muscle cells were present (figure 2c). Nuclei with significant gene activity for both fibroblast and myeloid markers were probably doublet nuclei and excluded from further analyses. Control fibroblasts clustered distinctly from the IPF fibroblasts (supplementary figure S4c). We focused our analysis on the comparison of IPF myofibroblasts to IPF nonmyogenic fibroblasts (including the adventitial, alveolar and *CXCL2*^{hi} fibroblast populations), as well as IPF to control fibroblasts. There were 163 DARs more accessible in IPF myofibroblasts *versus* 88 DARs more accessible in IPF nonmyogenic fibroblasts (by Bonferroni-adjusted $p<0.05$). Overall, these DARs were consistent across individual IPF samples (figure 2d). These DARs were annotated to the nearest gene and utilised for pathway analysis by IPA, with the stem-cell pluripotency, thioredoxin, hepatic fibrosis and regulation of epithelial–mesenchymal transition (EMT) among the top upregulated pathways in IPF myofibroblasts *versus* nonmyogenic fibroblasts (figure 2e). Of the DARs more accessible in IPF myofibroblasts, 17 were annotated to DEGs when comparing IPF myofibroblasts to nonmyogenic fibroblasts, including *SPON2*, *PLEKHG1*, *SMYD3*, *AKAP7*, *LOXL2* and *TSPAN2*. Only 19 DARs were more accessible in IPF fibroblasts compared to 25 DARS in control fibroblasts, revealing pathways not as clearly associated with fibrosis (figure 2f). In comparing IPF to control fibroblasts, five out of the 19 significant DARs were annotated to genes upregulated in IPF fibroblasts including *FBXL7*, *SPON2*, *ATP10D* and *RUNX1*. While traditionally annotated by least base pairs distance, *cis*-regulatory regions do not inevitably regulate the nearest gene, but may instead associate with more distant genes *via* three-dimensional chromatin looping. DARs not located near DEGs may regulate more distant genes *via* such long-range interactions.

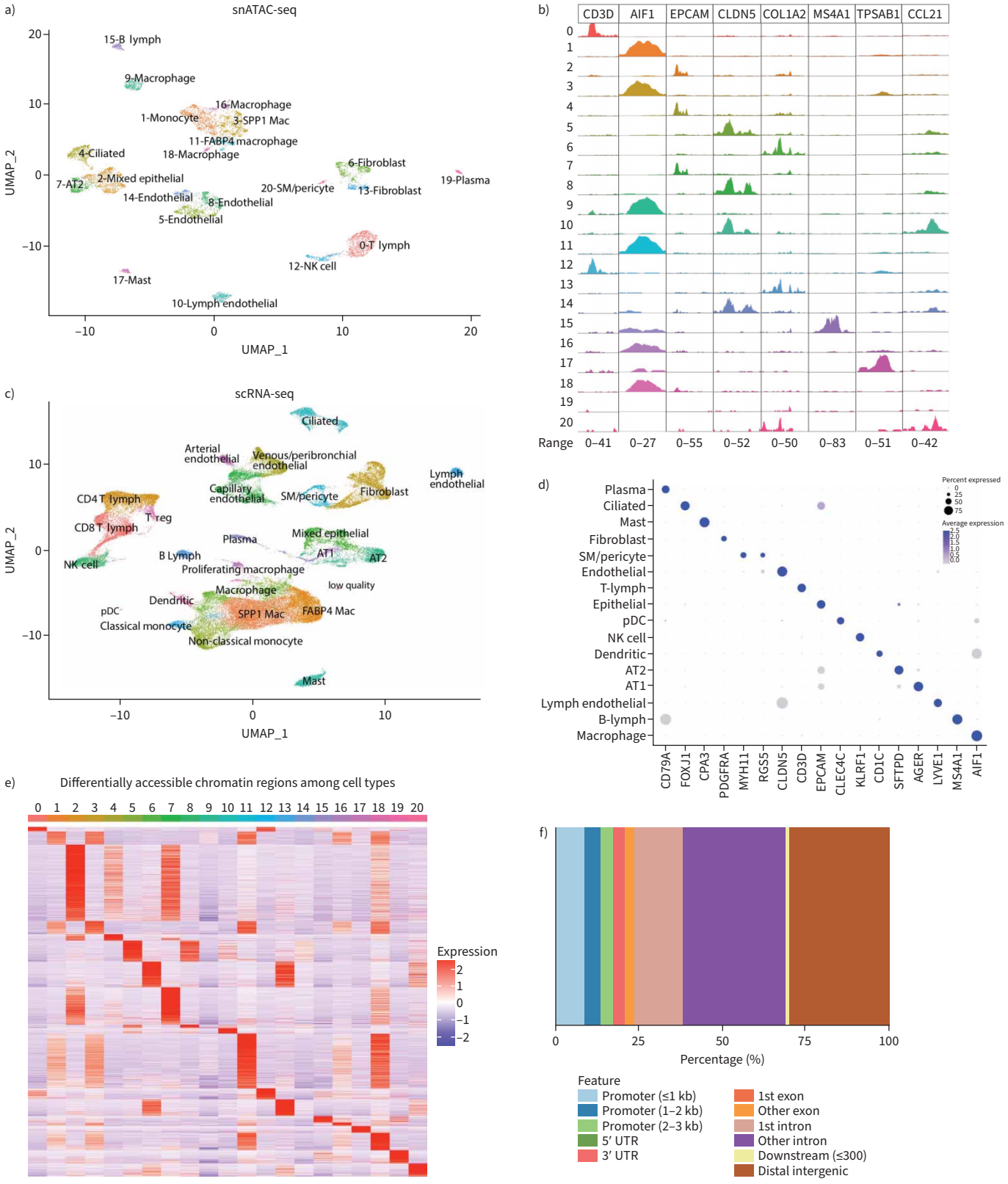


FIGURE 1 Single-cell RNA-sequencing (scRNA-seq) and single-nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq) profiling of the human idiopathic pulmonary fibrosis (IPF) and healthy lung. a) Uniform manifold approximation and projection (UMAP) plot snATAC-seq dataset (IPF n=3, control n=2) identified by cluster number and cell type. b) Merged coverage plots demonstrating pseudo-bulk chromatin accessibility (fragment coverage by frequency of Tn5 insertion) around marker gene promoters. Y-axis cluster numbers correspond to cell clusters

in figure 1a. Range of normalised accessibility for fragment coverage of each gene listed on x-axis. c) UMAP plot of scRNA-seq dataset (IPF n=10, control n=8) identified by cell type. d) Dot plot of scRNA-seq dataset showing gene expression of selected cell-type specific marker genes. The diameter of the dot corresponds to the proportion of the cells expressing the gene, and the colour density of the dot corresponds to the average expression level relative to all cell types. e) Heatmap of average number of Tn5 cut sites within the differentially accessible regions (DARs) (each row is a unique DAR) for each cell type. The colour scale represents a z-score of the number of Tn5 sites within each DAR. f) Genomic annotation of differentially accessible region locations. Mac: macrophage; AT: alveolar type; SM: smooth muscle; NK: natural killer; pDC: plasmacytoid dendritic cell; UTR: untranslated region.

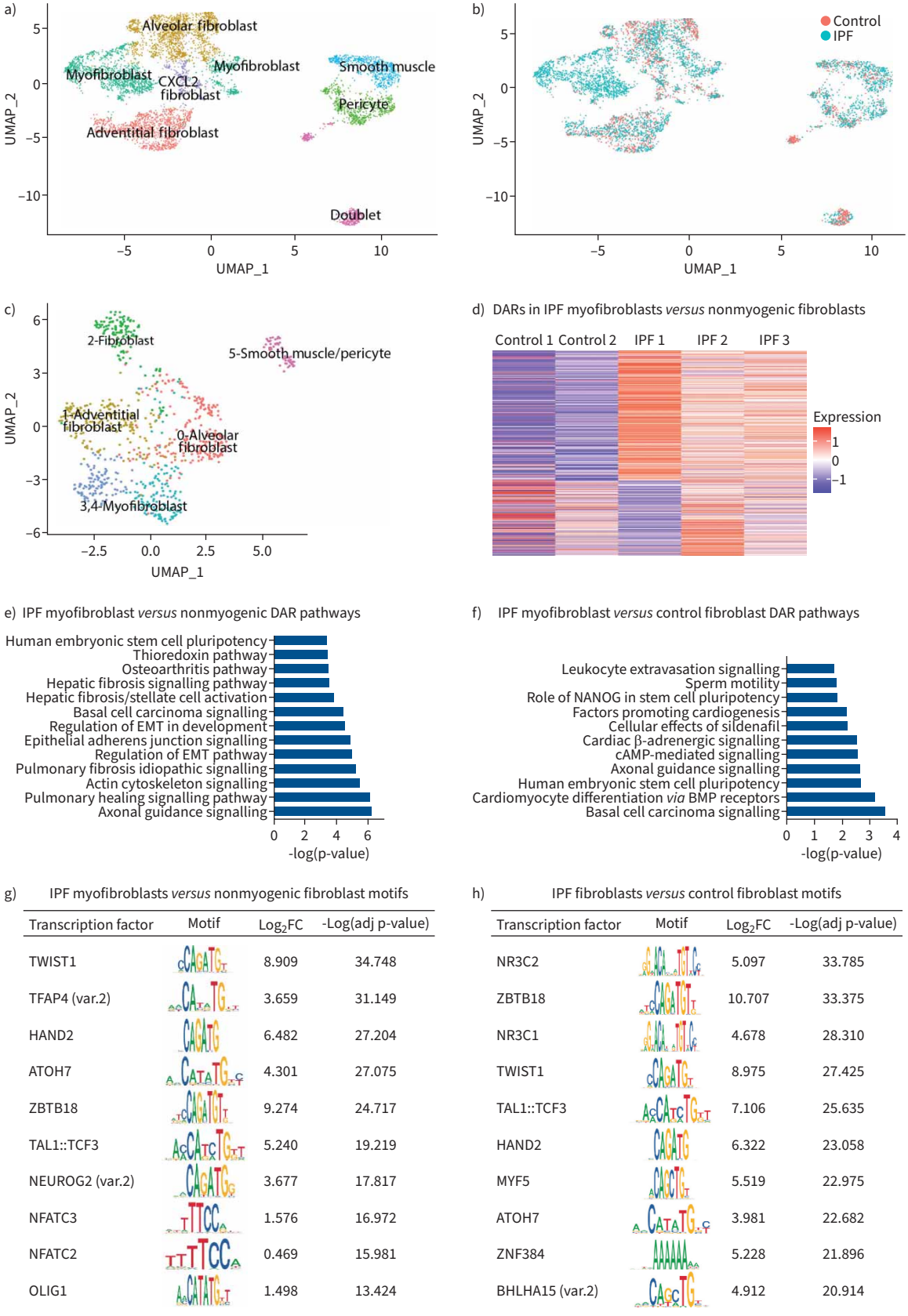
TWIST1 motif activity enrichment in IPF myofibroblasts

Transcription factor motif enrichment can be inferred for cell populations based on the enriched presence of transcription factor-binding motifs within accessible chromatin regions, predicting critical, active transcription factors regulating the cell state of interest. To assess transcription factor-motif activity we used chromVAR to determine transcription factor-associated accessibility in our snATAC-seq dataset [21]. Specific motifs were associated with each individual cell type, with known cell-type enriched transcription factors validating our data and analysis, such as FOXA1 and TEAD1 in alveolar epithelial cells [22], ETS1 in natural killer cells and T-lymphocytes [23] and MEF2C in smooth muscle cells [24]. Motifs highly enriched in fibroblasts compared to other cell types included ZBTB26, NFATC2, TWIST1, MYF5, HSF1, HSF2 and PBX2, among others (supplementary table S1, supplementary figure S6a).

On a global scale, transcription factor expression had limited correlation with transcription factor activity, supporting the notion that transcription factors may act as activators or repressors of gene expression based on post-transcriptional regulation of their activity. Of the 208 significant unique transcription factor motif activities identified when comparing IPF to control fibroblasts, only 36 correlated to DEGs between the two populations. Comparing IPF myofibroblasts to nonmyogenic fibroblasts, motifs for TWIST1, TFAP4, HAND2, ATOH7 and ZBTB18 were the most significantly enriched (figure 2g), with *TWIST1*, *TCF3* and *NFATC3* expressed more highly by myofibroblasts. To evaluate for bias by individual sample, a leave-one-out analysis was performed for motif enrichment, with *TWIST1* consistently noted among the top motifs in each scenario (supplementary material). Motif activity for these transcription factors was also consistent visually across samples (supplementary figure S7a). Comparing IPF to control fibroblasts, motifs for NR3C2, ZBTB18, NR3C1, TWIST1 and TAL1::TCF3 were the most significantly enriched, with *NR3C1* conversely having decreased expression in IPF fibroblasts (figure 2h).

Multiple transcription factors with a shared consensus binding sequence may bind to a highly similar motif, distinguished by minor differences in their position weight matrices, such as the E-box motifs of *TWIST1* and *HAND2*. *In silico* motif enrichment analyses alone cannot definitively determine which transcription factor binds to a particular motif. To confirm that TWIST1 binds to relevant regions of accessible chromatin in IPF fibroblasts, we performed chromatin immunoprecipitation-sequencing of pulmonary fibroblasts from patients with IPF (n=4) following TWIST1 antibody (Ab) immunoprecipitation and compared results to the snATAC-seq results. After quality control, pre-processing, peak calling utilising input controls and comparing peaks across the IPF samples, 66 statistically significant peaks occurred in multiple IPF TWIST1 Ab samples. 48% (32 out of 66) of overlapping peaks occurred in at least three of the four IPF samples (supplementary figure S7c), with "regulation of glutamatergic transmission" the most significantly enriched pathway among the genes annotated to these peaks. Despite the inevitable changes in transcription factor-biding and chromatin structure occurring in culture, 32% of the shared IPF TWIST1 Ab peaks occurred in regions of accessible peaks in the snATAC-seq fibroblast data. In particular, TWIST1 bound in the coding sequence of *KANK3* ($\log_2$ fold change (FC) 0.978, p=0.0539) and *DYNCH11* ($\log_2$FC 1.548, p=0.0166) in DARs between the IPF and control fibroblasts (figure 3). E-box motifs were present in these regions in snATAC-seq fibroblasts. TWIST1 also bound in DARs between IPF myofibroblast and nonmyogenic fibroblasts in regions annotated to *CACNG8*, *LINC00415*, *RP11-34F13.3* and *GPR27*.

As TWIST1 motif activity was enriched in IPF myofibroblasts and IPF fibroblasts and its expression is highly specific to myofibroblasts (figure 4a–c), we further investigated the cell populations showing increased TWIST1-associated DARs. We separated the IPF myofibroblasts into top-quartile TWIST1 motif activity, and those with low TWIST1 motif activity (bottom three quartiles). The thioredoxin pathway, calcium signalling, hepatic fibrosis and 3-phosphoinositide degradation were among the pathways significantly enriched by genes annotated to DARs in TWIST1^{hi} motif activity myofibroblasts (supplementary figure S7b). In addition to TWIST1, other E-box motif transcription factors including HAND2, ZBTB18, NEUROG2 and NEUROD1 had the highest motif enrichment in the TWIST1^{hi} motif activity myofibroblasts compared to those with TWIST1^{lo} motif activity.



g) IPF myofibroblasts *versus* nonmyogenic fibroblast motifs

Transcription factor	Motif	Log_2FC	-Log(adj p-value)
TWIST1		8.909	34.748
TFAP4 (var.2)		3.659	31.149
HAND2		6.482	27.204
ATOH7		4.301	27.075
ZBTB18		9.274	24.717
TAL1::TCF3		5.240	19.219
NEUROG2 (var.2)		3.677	17.817
NFATC3		1.576	16.972
NFATC2		0.469	15.981
OLIG1		1.498	13.424

h) IPF fibroblasts *versus* control fibroblast motifs

Transcription factor	Motif	Log_2FC	-Log(adj p-value)
NR3C2		5.097	33.785
ZBTB18		10.707	33.375
NR3C1		4.678	28.310
TWIST1		8.975	27.425
TAL1::TCF3		7.106	25.635
HAND2		6.322	23.058
MYF5		5.519	22.975
ATOH7		3.981	22.682
ZNF384		5.228	21.896
BHLHA15 (var.2)		4.912	20.914

FIGURE 2 Fibroblast subpopulations and transcription factor motif activity in idiopathic pulmonary fibrosis (IPF) myofibroblasts. a) Uniform manifold approximation and projection (UMAP) plot of single-cell RNA-sequencing (scRNA-seq) fibroblasts, smooth muscle cells and pericyte

clusters by cell identity. b) UMAP plot of scRNAseq fibroblasts, smooth muscle cells and pericyte clusters from (a) with cells depicted by origination from IPF *versus* control samples. c) UMAP plot of single-nucleus assay for transposase-accessible chromatin sequencing fibroblast, smooth muscle cells and pericyte clusters by cell identity. d) Heatmap of average number of Tn5 cut sites within the differentially accessible regions when comparing IPF myofibroblasts to IPF nonmyogenic fibroblasts, depicted by individual sample. Each row is a unique differentially accessible region (DAR). e) Ingenuity pathway analysis pathways significantly enriched for genes annotated to the upregulated DARs in IPF myofibroblasts *versus* IPF nonmyogenic fibroblasts. f) Ingenuity pathway analysis pathways significantly enriched for genes annotated to the upregulated DAR in IPF fibroblasts *versus* control fibroblasts. g) Transcription factors with the most significantly enriched motif activity when comparing IPF myofibroblasts to the IPF nonmyogenic fibroblasts. h) Transcription factors with the most significantly enriched motif activity when comparing all IPF fibroblasts to all control fibroblasts. EMT: epithelial–mesenchymal transition; BMP: bone morphogenetic protein; FC: fold change; adj: adjusted.

TWIST1 expression was upregulated in IPF myofibroblasts (log_2FC 3.136, adjusted $p=1.41\times10^{-24}$) *versus* nonmyogenic fibroblasts. Compared to other E-box motif transcription factors, TWIST1 was the most specific to myofibroblasts and showed the greatest upregulation in IPF (figure 4f), supporting our hypothesis that *TWIST1* dysregulation perpetuates IPF myofibroblast activity. In comparing *TWIST1*^{hi} IPF fibroblasts (expression level >0.5, top 6.5% expression) to *TWIST1*^{lo} IPF fibroblasts (expression level <0.5), we identified 338 DEGs including upregulation of *POSTN*, *ASPN*, *LOXL2*, *MMP14* and *COL8A1*, among others, in the *TWIST1*^{hi} population. Enriched pathways in *TWIST1*^{hi} IPF fibroblasts included pulmonary fibrosis idiopathic signalling, hepatic fibrosis, axonal guidance signalling and S100 family signalling.

Two regions of *TWIST1*, in the promoter and the second intron, had significant differential accessibility in comparing myofibroblast to nonmyogenic fibroblasts (figure 4d). In addition, we also examined myofibroblasts with accessible chromatin in *TWIST1* (51.2% of myofibroblasts) *versus* those with closed chromatin in *TWIST1* (48.8% of myofibroblasts). Pathways enriched among genes annotated to DARs in *TWIST1*-open myofibroblasts included fibroblast growth factor (FGF) signalling, regulation of EMT, Gαq signalling and ribonucleotide reductase signalling, among others (figure 4e). Regulation of EMT was enriched based on enhanced accessibility for *CD70*, *FGF13*, *FGFR1*, *MAPK3*, *PARD6B*, *SNAI2*, *STAT3* and *TWIST1*, supporting the previously identified central role of TWIST1 in EMT [25], as well as supporting the validity of our data.

To investigate whether TWIST1 motifs were enriched in nonmesenchymal cells known to transition to mesenchymal and myofibroblast phenotypes in fibrosis (*via* EMT or endothelial mesenchymal transition) we evaluated enriched motifs in IPF *versus* control epithelial and endothelial populations. TWIST1 motifs were significantly enriched in IPF endothelial nuclei *versus* controls; however, the nuclear factor of activated T-cell-related factors (NFATC4, NFATC3 and NFATC2), several AP-1 family transcription factors and NR3C1 were the top enriched motifs for this comparison (supplementary table S6). TWIST1 motifs were not significantly enriched in the aggregate IPF epithelial cells *versus* controls, or in the comparison for alveolar type 1, alveolar type 2, basal or ciliated cells; however, the more limited number of nuclei for these populations limited the number of significant motifs identified. These data support the primary action of TWIST1 within the myofibroblast, rather than the epithelial populations, consistent with its increased gene expression in the myofibroblast population.

Increased expression of Twist1 *in collagen-producing cells is associated with increased collagen synthetic activity*

We have observed previously that IPF patients with the highest expression of *TWIST1* by whole-lung microarray analysis exhibit the most impaired gas exchange [26]. Combining these data with our ATAC-seq observations led us to consider how *Twist1* overexpression in fibroblasts may impact an animal model of pulmonary fibrosis. We examined the effect of induced expression of *Twist1* in lung *Col1a2*$^{+}$ expressing fibroblasts (Twist1-LUC; figure 5a). Lung fibroblasts were isolated and cultured in the presence of tamoxifen and doxycycline, to induce *Twist1* expression, with and without transforming growth factor (TGF)-β. In unstimulated Twist1-LUC fibroblasts, we observed increased expression of collagen I and the myofibroblast marker α-smooth muscle actin (α-SMA/ACTA2) compared to wild-type fibroblasts (Twist1-WT) (figure 5b–d). TGF-β augmented both α-SMA/ACTA2 and collagen 1 in the presence of *Twist1* overexpression, suggesting that TWIST1 mediates a TGF-β independent pathway. We confirmed increased expression of *Twist1* in Twist1-Luc fibroblasts (figure 5b and e). Although *Twist1* expression may promote a "pro-survival" phenotype in certain tumours, we found that *Twist1* overexpression did not promote proliferation or resistance to apoptosis by pulmonary fibroblasts *in vitro* (supplementary figure S8a–c).

Next, we explored the effect of increased *Twist1* in fibroblasts *in vivo* following bleomycin injury (figure 5f–g, supplementary figure S9a). Uninjured Twist1-Luc mice showed no pathology (supplementary figure S9b).

Following bleomycin injury, we saw no difference in the ratio of wet-to-dry lung mass, suggesting a comparable degree of acute lung injury between genotypes (figure 5f). In contrast, we observed increased collagen content in Twist1-LUC mice injured with bleomycin compared to Twist1-WT mice (figure 5g). Histologically, we observed comparable acute lung injury in both Twist1-WT and Twist1-LUC mice

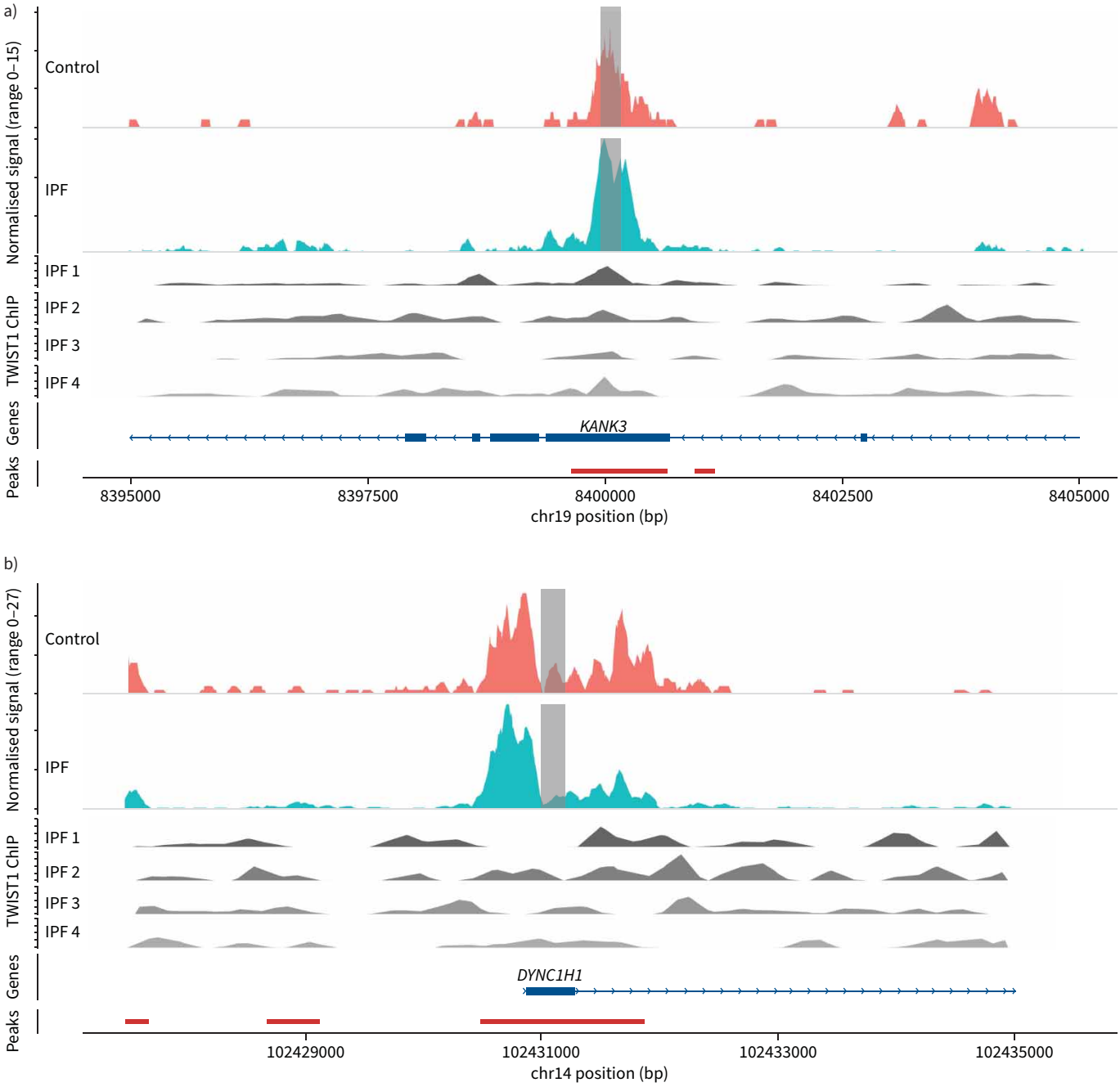


FIGURE 3 TWIST1 binds in areas of *in vivo* accessible chromatin in idiopathic pulmonary fibrosis (IPF) fibroblasts. a) Coverage plot demonstrating Tn5 insertion frequency (pseudobulk single-nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq) tracks) in IPF and control fibroblasts in the region of KANK3 with ATAC-seq peaks depicted by red bars. TWIST1 Ab chromatin immunoprecipitation sequencing (ChIP-seq) peaks from IPF explant fibroblasts (grey tracks labelled IPF 1–4) in the same region are depicted below. Grey shading indicates the region of TWIST1 binding by ChIP-seq experiments. Scales for ATAC-seq tracks and ChIP-seq tracks are independent. b) Coverage plot demonstrating Tn5 insertion frequency (pseudobulk ATAC-seq tracks) in IPF and control fibroblasts in the region of DYNC1H1 with ATAC-seq peaks depicted by red bars. TWIST1 Ab ChIP-seq peaks from IPF explant fibroblasts (grey tracks labelled IPF 1–4) in the same region are depicted below. Light grey box indicates the region of TWIST1 binding by ChIP-seq experiments. Scales for ATAC-seq tracks and ChIP-seq tracks are independent.

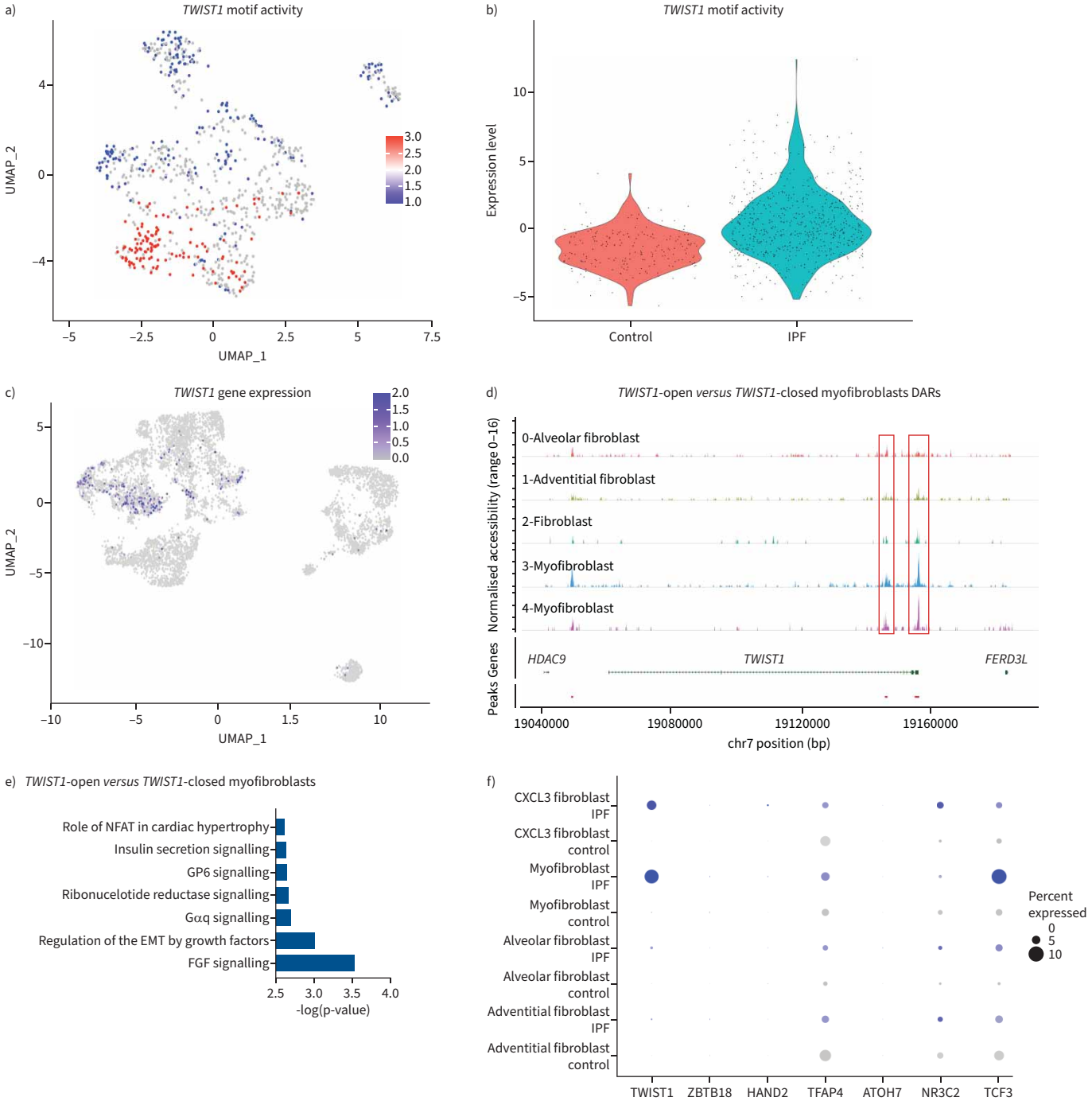


FIGURE 4 TWIST1 expression and motif activity in idiopathic pulmonary fibrosis (IPF) and control fibroblasts. a) Uniform manifold approximation and projection (UMAP) plot of TWIST1 motif activity in the mesenchymal single-nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq) clustering, depicted by scaled expression with red representing the highest motif activity. b) Violin plot of TWIST1 motif activity comparing IPF *versus* control fibroblasts only. c) UMAP plot of TWIST1 gene expression in the mesenchymal single-cell RNA-sequencing (scRNA-seq) clustering demonstrating high TWIST1 expression in the myofibroblasts only. d) Coverage plot demonstrating Tn5 insertion frequency by snATAC-seq fibroblast cluster in the TWIST1 gene region. Red boxes represent regions of statistically significant differential accessibility in IPF myofibroblasts *versus* nonmyogenic fibroblasts. e) Ingenuity pathway analysis pathways significantly enriched for genes annotated to the upregulated differentially accessible regions (DAR) in TWIST1-open chromatin myofibroblasts *versus* TWIST1-closed chromatin myofibroblasts. f) Dot plot of E-box transcription factor gene expression in IPF and control mesenchymal populations by scRNA-seq with deeper colour indicating higher level of gene expression and circle size indicating the percentage of cells in the population expressing the gene. NFAT: nuclear factor of activated T-cells; EMT: epithelial–mesenchymal transition; FGF: fibroblast growth factor.

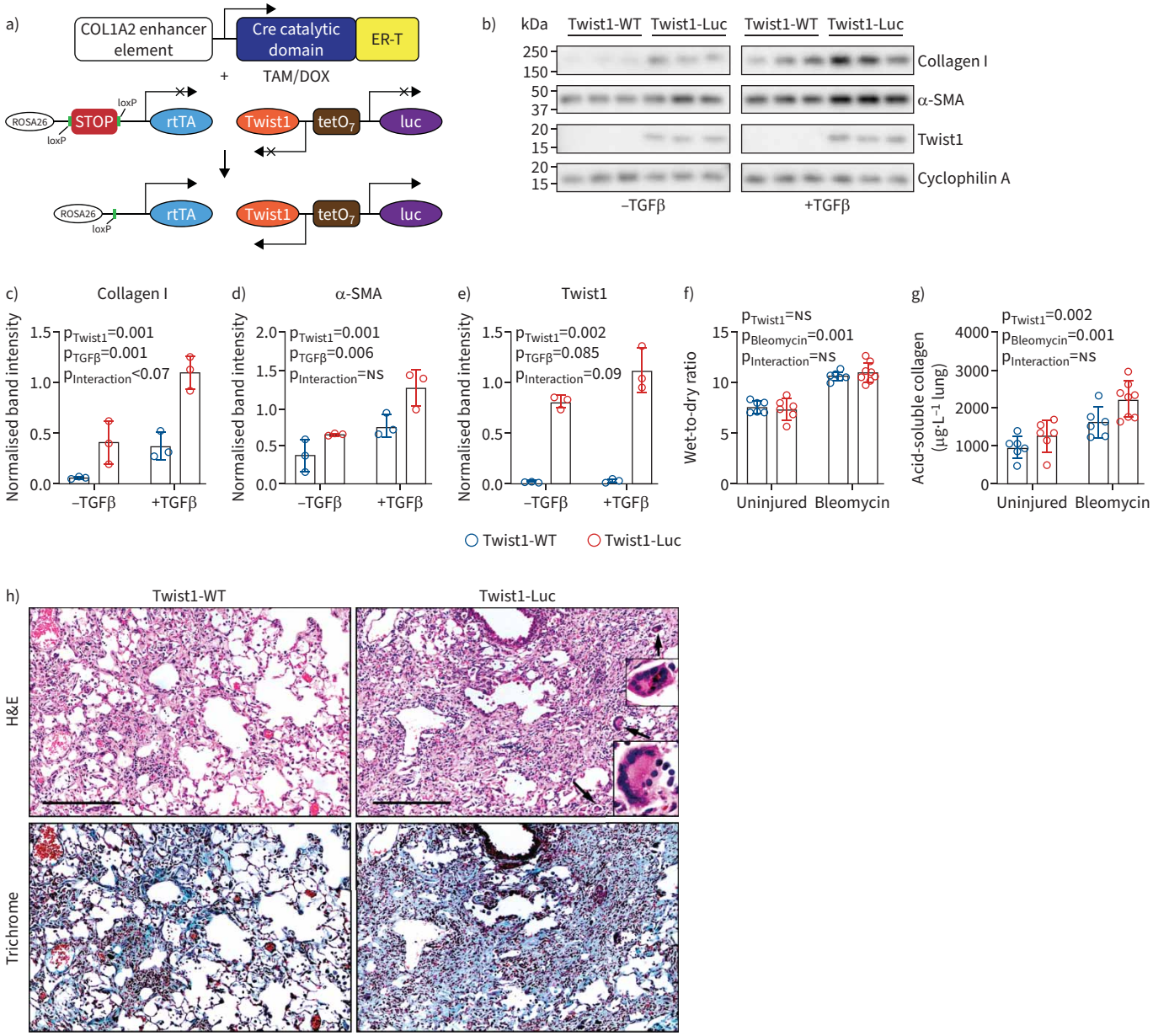


FIGURE 5 Overexpression of Twist1 in Col1a2^{+} cells leads to increased collagen I levels *in vitro* and *in vivo*. a) A triple transgenic animal was bred where Cre recombinase is under control of the col1a2 enhancer element (col1a2-Cre-ER(T)). In the presence of tamoxifen (TAM), the STOP signal is excised leading to expression of the reverse tetracycline transactivator (rtTA). In the presence of doxycycline (DOX) and the rtTA, the tetO7 operator is activated leading to expression of Twist1 and luciferase. b) Lung fibroblasts from Twist1-WT (wild-type) and Twist1-Luc (Twist1 overexpressors) were incubated in the presence of TAM and DOX with and without transforming growth factor (TGF)-β (2 ng·mL^{-1}). Cells were lysed and subjected to immunoblotting for collagen I, α-smooth muscle actin (SMA), Twist1 and the loading control, cyclophilin A. In the presence of TAM/DOX, increased c) collagen, d) α-SMA and e) Twist1 in Twist1-LUC fibroblasts. This was amplified in the presence of TGF-β (n=3). Data were analysed by robust nonparametric two-way ANOVA. p-values for the effects of Twist1 and TGF-β and the interaction are presented in the panels. f) Twist1-WT and Twist1-Luc mice were injured with bleomycin. Animals were sacrificed at 14 days. Lungs were excised, and the ratio of the lung mass before and after freeze-drying was determined. g) Determination of acid-soluble collagen content showed a significant increase in bleomycin-induced collagen in Twist1-Luc mice compared to Twist1-WT mice (by robust nonparametric two-way ANOVA; n=6–8). h) A comparable degree of histological injury was observed in Twist1-WT and Twist1-Luc mice. Haematoxylin and eosin (H&E) and trichrome images are presented. Incidentally noted multinucleated giant cells are identified and magnified by the black arrows. Scale bar=200 μm, inset ×100 magnification. NS: nonsignificant.

injured with bleomycin, but increased collagen deposition in Twist1-LUC mice (figure 5h). We incidentally noted airspace multinucleated giant cells in bleomycin-treated Twist1-LUC mice. As expected, *Twist1* mRNA was overexpressed in whole lungs of Twist1-LUC mice (supplementary figure S10a).

Bleomycin induced the inflammatory mediators *Tnfa*, *Il1*, *Il6*, *Cxcl12 and Ccl7*; however, *Twist1* overexpression in Col1a2^{+} cells did not alter expression of these mediators except for *Cxcl12* and *Il6*, which were slightly decreased (supplementary figure S10a–f). Taken together, these data show that increased expression of *Twist1* in collagen-producing cells is associated with increased collagen synthetic activity in both *in vitro* and *in vivo* models.

Timing of Twist1 *expression*

To investigate the timing of *Twist1* expression in lung injury and fibrosis, we analysed publicly available whole-lung scRNA-seq data collected at days 3, 7, 10, 14, 21 and 28 of the murine intratracheal bleomycin injury model [27]. Similar to our human IPF scRNA-seq studies, *Twist1* was primarily expressed in myofibroblasts (supplementary figure S11a and b). Within the mesenchymal compartment, no *Twist1* expression was detected in the uninjured control, day 3 or day 7 animals. *Twist1* expression was present in the highest number of cells at day 10 of bleomycin injury, with a lower percentage of cells expressing it at days 14, 21 and 28 (supplementary figure S11c). Day 10 *Twist1* expression correlates with early extracellular matrix deposition in the bleomycin injury model (typically peaking at day 14) [28], as well as peak alveolar *Krt8* expression, reflecting a transitional alveolar cell present in lung injury that persists within the fibrotic lung [27].

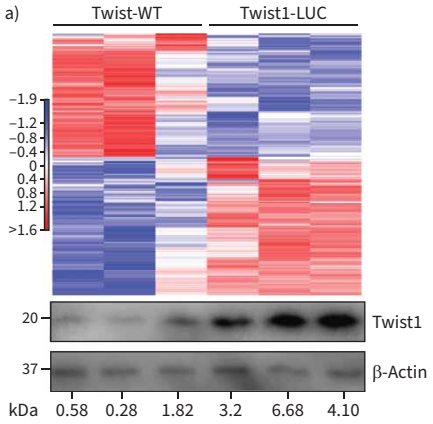
Twist1 *overexpression induces dysregulation of multiple profibrotic genes*

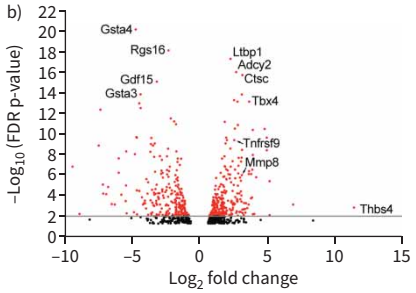
To further explore genes regulated by TWIST1, we compared gene expression in cultured lung fibroblasts from three Twist1-LUC mice and three Twist1-WT mice *in vitro* by RNA-seq (figure 6). On clustering by DEGs between Twist1-LUC and Twist1-WT fibroblasts, one of the three Twist1-WT fibroblasts clustered more closely to Twist1-LUC fibroblasts. By immunoblotting, this line spontaneously expressed higher TWIST1 protein than the other two Twist1-WT lines, but less than the Twist1-LUC fibroblasts, thus representing an intermediate phenotype (figure 6a). Several significantly upregulated genes (figure 6b–d) have been associated with TGF-β signalling and pulmonary fibrosis including *Ltbp1*, *Tbx*, *Tnc* and *Thbs4*. Based on IPA dysregulated canonical pathways including "systemic lupus erythematosus in B cell signalling" (*Tnfsf11* and *Tnfsf15*) and "role of hypercytokinemia in the pathogenesis of influenza" as well as Hippo signalling (implicated in pulmonary fibrosis [29]) and pulmonary and hepatic fibrosis pathways (*Ptch2*, *Il1rap*, *Itgb3* and *Flt3*). Downregulated signalling for xenobiotic metabolism, glutathione-mediated detoxification and NRF2-mediated oxidative stress response (*Gsta3*, *Nqo1*, *Cyp1a1* and *Acta1*) support previous data suggesting that loss of these pathways is a component of fibrosis (figure 6e) [30, 31]. In the heatmap-based on genes showing the highest coefficient of variation, fibrotic genes such as *Col1a1*, *Col1a2*, *Col3a1*, *Tnc*, *Thbs1* and *Lox* were highly upregulated in Twist1-LUC fibroblasts (supplementary figure S12). Using quantitative reverse-transcriptase PCR, we validated the top-most DEGs genes that are upregulated and downregulated, as well as extracellular matrix genes (supplementary figure S10). In comparing our snATAC-seq data with the mouse lung fibroblasts, we observed that *MMP8* and *TNFRSF9*, among the top upregulated genes in Twist1-LUC fibroblasts (figure 6c), have significantly increased chromatin accessibility in IPF myofibroblasts and a trend towards increased accessibility in IPF *versus* control fibroblasts (supplementary figures S13 and S14). This finding demonstrates that downstream targets of TWIST1 have altered chromatin in myofibroblasts.

Discussion

Our study demonstrates that the differentiation of myofibroblasts – the central effector cells in IPF – is characterised by a significant shift in chromatin accessibility, dominated by opening of E-box transcription factor binding sites. We utilised single-cell sequencing platforms of *ex vivo* IPF lungs to obviate the distortion of signals across heterogeneous populations and the changes in chromatin accessibility and gene expression that may occur with expansion in culture. Of the E-box transcription factors, we identified TWIST1 as the most highly enriched regulator of myofibroblast activity. We then confirmed a critical regulatory role for TWIST1 by demonstrating that overexpression of Twist1 in the fibroblast compartment, *in vitro* and *in vivo*, led to increased expression of collagen I and α-SMA.

Previous studies have identified epigenetic changes in IPF lungs *via* methylation profiling [32–34]; however, knowledge of cell-type specific epigenetic alterations remains limited. A recent study by Hanmandlu *et al.* [35] utilised bulk ATAC-seq of cultured fibroblasts to investigate chromatin accessibility in IPF upper-lobe fibroblasts. They similarly identified enrichment of the E-box transcription factors TWIST1 and ZBTB18 motifs in IPF fibroblasts, further supporting an important role for TWIST1 and the E-box transcription factors in IPF myofibroblasts. However, other motifs implicated by their analysis, including FOXA1 and FOXP1, were significantly less enriched in our analyses, while CBFB was enriched in IPF fibroblasts in the bulk analyses only, indicating that accessibility may be altered by *in vitro* culture, though heterogeneity among samples cannot be ruled out. The E-box transcription factor MYF5, a known regulatory factor critical to myogenic differentiation, was also enriched in our snATAC-seq, but not the cultured fibroblasts.





c)

Name	Log_2 fold change	$-Log_{10}$ (FDR p-value)
Ltbp1	2.3	17.3
Adcy2	2.7	16.0
Ctsc	3.2	15.7
Itga8	2.5	13.3
Tbx4	3.7	13.2
Masp1	2.8	13.1
Htr2a	1.9	11.2
Ccdc85a	2.6	10.5
Chl1	4.8	10.5
Ptch2	3.9	10.4
Il1rap	2.0	9.7
Mdga2	5.0	9.7
Tnfrsf9	2.6	9.4
Ccbe1	2.0	9.4
Itgbl1	1.9	9.1
Nr4a3	2.3	8.6
Acta1	3.1	8.5
Angpt1	5.0	8.4
Cpxm2	3.9	8.0
C1qtnf5	1.8	7.8
Clca3a1	3.1	7.7
Serpina3g	2.9	7.7
Mmp8	3.1	6.0
Serpina3n	2.2	2.7
Tnfsf11	3.3	1.7
Thbs1	1.3	1.7
Tnc	1.3	1.5

d)

Name	Log_2 fold change	$-Log_{10}$ (FDR p-value)
Gsta4	−4.7	20.2
Rgs16	−2.3	18.1
Gdf15	−3.2	15.1
Gsta3	−4.3	13.9
Des	−4.4	13.0
Nqo1	−4.3	12.5
Gsta1	−7.3	12.4
Slc6a17	−2.1	11.5
Cotl1	−1.9	11.2
Mustn1	−1.7	11.0
Megf10	−4.6	9.7
Trdc	−4.6	9.7
Serpinb6c	−3.6	9.6
Abcc4	−3.3	9.1
Slc7a11	−3.6	9.0
Cyp1a1	−7.4	8.9
Mmd	−3.4	8.9
Arhgap20	−2.3	8.6
Ecel1	−5.4	8.3
Olfm1	−1.7	8.3
Star	−3.6	8.3
Gm8797	−4.8	8.0
Enpp2	−2.9	7.9
Cst6	−2.9	7.9
Itgb3	−2.4	7.8
Upk1b	−3.3	7.8
Trim12a	−6.0	7.6

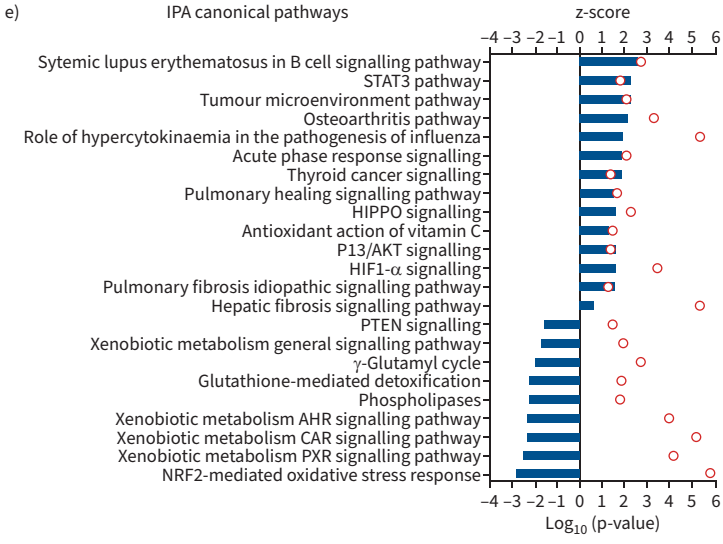


FIGURE 6 Twist1 overexpression in mouse lung fibroblasts is associated with dysregulation of several pulmonary fibrosis genes and pathways. Bulk RNA-sequencing (seq) was performed on fibroblasts isolated

from lungs of wild-type (WT) and Twist1-LUC mice (n=3). Estimation of differential gene expression using CLC Genomics Workbench was performed comparing fibroblasts from knock-in with normal lungs. a) Hierarchical clustering heatmap of significant differentially expressed genes was generated using CLC Genomics Workbench using minimum absolute fold change of 3.0 and false discovery rate (FDR) p-value threshold of 0.05. Immunoblotting is shown for the individual lines subjected to RNA-seq. Densitometry normalised to β-actin is shown beneath. b) Volcano plot shows comparative analysis of differentially expressed genes between the WT and Twist1-LUC. c) List of dysregulated genes by FDR that are upregulated ranked by $-\log_{10}$ (FDR p-value) with p<0.05 cut-off. d) List of top downregulated genes ranked by $-\log_{10}$ (FDR p-value) with p<0.05 cut-off downregulated in Twist1-LUC fibroblasts compared to Twist1-WT fibroblasts. e) Ingenuity pathway analysis (IPA) of dysregulated canonical pathways between WT and Twist1-LUC (n=3) by z-score and FDR.

The consistent enrichment of E-box motifs in IPF myofibroblasts suggests modulation of their accessibility as a critical step within or resulting from activation of the aberrant myofibroblast programme. While we focus in this work on TWIST1, the complex mechanism of myofibroblast activation undoubtedly involves the coordinated activation of multiple transcription factors. Defining such shifts in a cell population's epigenetic state now opens the door for novel molecular and computational approaches to therapeutic development in IPF. For instance, the rapid advancement of small-molecule therapeutics including those inhibiting DNA–protein binding and transcription factor complexes supports identifying and targeting key dimerisation pairs or other coordinated binding partners of TWIST1 [36]. Investigating chromatin alterations in the context of therapeutics may identify agents halting (and ideally reversing) such epigenetic changes, potentially as an early signal for therapeutic response.

In contrast to previous studies [26, 37–39] examining TWIST1 in fibroblasts, to our knowledge, we are the first to report that fibroblast-specific overexpression of *Twist1 in vivo* is associated with increased collagen synthesis following bleomycin injury. This corroborates our observation that increased expression of *TWIST1* in IPF is associated with worse gas exchange [26]. It was striking that one Twist1-WT line of fibroblasts, through experimental variation, mapped more closely with the Twist1-LUC fibroblasts, suggesting a very narrow dynamic range of expression of TWIST1 in unstimulated cells. Deviations from that narrow range of expression can lead to pronounced differences in fibroblast phenotypes [26]. Taken together, these data support a unique model whereby TWIST1 serves as a critical "rheostat" in IPF. Cellular levels of TWIST1 are tightly regulated, and even small changes can significantly impact the fibrotic phenotype. It is clear that the "good/bad" paradigm does not suffice in describing the role of TWIST1 in pulmonary fibrosis. Downstream studies which will require much greater depth include 1) what are TWIST1's binding partners? [40] and 2) do other E-box transcription factors compensate for the loss of TWIST1?

Our studies have focused specifically on the behaviour of *Twist1* expression in fibroblasts. In our hands, *Twist1* expression did not protect fibroblasts from apoptosis *in vitro*. An important remaining question would be the persistence of Twist1-LUC fibroblasts as fibrosis "resolves" after bleomycin injury in mice [41, 42]. In addition, we did not fully assess the effect of *Twist1* overexpression on other cell lineages in the lung as well as the more complex question of modelling human pulmonary fibrosis. This is clearly important, as we, surprisingly, observed multinucleated giant cells (MGCs) in bleomycin-injured Twist1-LUC mice. This phenotype in macrophages may be driven by increased expression of receptor activator of NF-κB ligand (RANKL) signalling (essential for MGC formation in bone [43]) in mouse lung fibroblasts following overexpression of *Twist1*. This finding suggests that *Twist1* overexpression impacts other cell types in the lung and may have important effects on epithelial cell function [40, 44]. Future studies should include testing Twist1-LUC mice in other models of pulmonary fibrosis and would include opportunities for single-cell transcriptomics.

As the demonstrated changes in *TWIST1* motif enrichment were consistently observed despite the modest sample size and TWIST1 binds in accessible chromatin in IPF fibroblasts, it merited further mechanistic investigation. Despite small snATAC-seq sample numbers, we demonstrate consistency of DARs and motif activity between cell types and IPF and control fibroblasts across individual samples. All IPF samples were from patients with end-stage disease receiving care at a tertiary medical centre and may not reflect the comprehensive IPF population.

In summary, our analysis utilises human multiomic single-cell analyses combined with *in vivo* murine disease models to investigate transcription factor networks critical to IPF myofibroblasts. Comparison of *in vivo* IPF myofibroblasts to nonmyogenic and control fibroblasts identified a dynamic opening of E-box transcription factor binding sites, with the E-box transcription factor TWIST1 particularly implicated as a

positive regulator of myofibroblast activity. Both low [26] and high expression of *Twist1* in fibroblasts is associated with increased collagen deposition in the lung, confirming its role as a critical regulator in the fibrotic lung. Future studies delineating the global mechanism modulating E-box transcription factor motif accessibility may identify crucial therapeutic targets for deactivating the aberrant myofibroblast programme.

Acknowledgements: The authors would like to acknowledge the Center for Organ Recovery and Education (CORE) as well as organ donors and their families for the generous donation of tissues used in this study. CLC Genomics Workbench software licensed through the Molecular Biology Information Service of the Health Sciences Library System, University of Pittsburgh, was used for data analysis. This research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided. Specifically, this work used the HTC cluster, which is supported by NIH award number S10OD028483.

Author contributions: E. Valenzi designed research studies, conducted experiments, acquired data, analysed data and wrote the manuscript. H. Bahudhanapati conducted experiments, acquired data, analysed data and wrote the manuscript. J. Tan designed research studies, conducted experiments, acquired data, analysed data and wrote the manuscript. T. Tabib conducted experiments, acquired data and analysed data. D.I. Sullivan analysed data. M. Nouraie performed statistical analyses on *in vitro* and *in vivo* data. J. Sembrat conducted experiments and acquired data. L. Fan acquired data and provided reagents. K. Chen acquired data and provided reagents. S. Liu performed ChIP-seq data analysis. M. Rojas acquired data and provided samples. A. Lafargue acquired data and provided reagents. D.W. Felsher provided hybrid mouse and wrote the manuscript. P.T. Tran provided hybrid mouse and wrote the manuscript. D.J. Kass designed research studies, acquired data, analysed data and wrote the manuscript. R. Lafyatis designed research studies, analysed data and edited the manuscript.

Conflict of interest: R. Lafyatis reports the following conflicts of interest outside the scope of work of this manuscript: R. Lafyatis has served as a consultant for Pfizer, Bristol Myers Squibb, Boehringer Ingelheim, Formation, Sanofi, Boehringer-Mannheim, Merck and Genentech/Roche, and holds or recently had research grants from Corbus, Formation, Moderna, Regeneron, Pfizer and Kiniksa, and holds equity in Thirona. All other authors report no relevant conflicts of interest.

Support statement: Support for the studies was provided by R01 HL 126990 to D.J. Kass, P50 AR 060780-06A1 to R. Lafyatis and D.J. Kass, and from the Pulmonary Fibrosis Foundation and National Scleroderma Foundation to E. Valenzi.

References

1 Bagnato G, Harari S. Cellular interactions in the pathogenesis of interstitial lung diseases. *Eur Respir Rev* 2015; 24: 102–114.
2 Darby IA, Zakuan N, Billet F, *et al.* The myofibroblast, a key cell in normal and pathological tissue repair. *Cell Mol Life Sci* 2016; 73: 1145–1157.
3 Pardo A, Selman M. The interplay of the genetic architecture, aging, and environmental factors in the pathogenesis of idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2021; 64: 163–172.
4 Hinz B, Lagares D. Evasion of apoptosis by myofibroblasts: a hallmark of fibrotic diseases. *Nat Rev Rheumatol* 2020; 16: 11–31.
5 Morse C, Tabib T, Sembrat J, *et al.* Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J* 2019; 54: 1802441.
6 Adams TS, Schupp JC, Poli S, *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* 2020; 6: eaba1983.
7 Habermann AC, Gutierrez AJ, Bui LT, *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* 2020; 6: eaba1972.
8 Reyfman PA, Walter JM, Joshi N, *et al.* Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am J Respir Crit Care Med* 2019; 199: 1517–1536.
9 Rubin AJ, Parker KR, Satpathy AT, *et al.* Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* 2019; 176: 361–376.
10 Cusanovich DA, Daza R, Adey A, *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 2015; 348: 910–914.
11 Valenzi E, Tabib T, Papazoglou A, *et al*. Disparate interferon signaling and shared aberrant basaloid cells in single-cell profiling of idiopathic pulmonary fibrosis and systemic sclerosis-associated interstitial lung disease. *Front Immunol* 2021; 12: 595811.
12 Valenzi E, Bulik M, Tabib T, *et al*. Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. *Ann Rheum Dis* 2019; 78: 1379–1387.

13 Stuart T, Srivastava A, Madad S, *et al*. Multimodal single-cell chromatin analysis with Signac. *Nat Methods* 2021; 18: 1333–1341.
14 Korsunsky I, Millard N, Fan J, *et al*. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019; 16: 1289–1296.
15 Schep AN, Wu B, Buenrostro JD, *et al*. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* 2017; 14: 975–978.
16 Zhang Y, Liu T, Meyer CA, *et al*. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* 2008; 9: 958.
17 Kass D, Bridges RS, Borczuk A, *et al*. Methionine aminopeptidase-2 as a selective target of myofibroblasts in pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2007; 37: 193–201.
18 Stuart T, Butler A, Hoffman P, *et al*. Comprehensive integration of single-cell data. *Cell* 2019; 177: 1888–1902.e21.
19 Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 2015; 31: 2382–2383.
20 Travaglini KJ, Nabhan AN, Penland L, *et al*. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 2020; 587: 619–625.
21 Weirauch MT, Yang A, Albu M, *et al*. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 2014; 158: 1431–1443.
22 Zhou B, Stueve TR, Mihalakakos EA, *et al*. Comprehensive epigenomic profiling of human alveolar epithelial differentiation identifies key epigenetic states and transcription factor co-regulatory networks for maintenance of distal lung identity. *BMC Genomics* 2021; 22: L879.
23 Taveirne S, Wahlen S, Van Loocke W, *et al*. The transcription factor ETS1 is an important regulator of human NK cell development and terminal differentiation. *Blood* 2020; 136: 288–298.
24 Hirai H, Yang B, Garcia-Barrio MT, *et al*. Direct reprogramming of fibroblasts into smooth muscle-like cells with defined transcription factors—brief report. *Arterioscler Thromb Vasc Biol* 2018; 38: 2191–2197.
25 Ning X, Zhang K, Wu Q, *et al*. Emerging role of Twist1 in fibrotic diseases. *J Cell Mol Med* 2018; 22: 1383–1391.
26 Tan J, Tedrow JR, Nouraie M, *et al.* Loss of Twist1 in the mesenchymal compartment promotes increased fibrosis in experimental lung injury by enhanced expression of CXCL12. *J Immunol* 2017; 198: 2269–2285.
27 Strunz M, Simon LM, Ansari M, *et al.* Alveolar regeneration through a $Krt8^{+}$ transitional stem cell state that persists in human lung fibrosis. *Nat Commun* 2020; 11: 3559.
28 Tashiro J, Rubio GA, Limper AH, *et al.* Exploring animal models that resemble idiopathic pulmonary fibrosis. *Front Med* 2017; 4: 118.
29 Haak AJ, Kostallari E, Sicard D, *et al.* Selective YAP/TAZ inhibition in fibroblasts *via* dopamine receptor D1 agonism reverses fibrosis. *Sci Transl Med* 2019; 11: eaau6296.
30 Hecker L, Logsdon NJ, Kurundkar D, *et al.* Reversal of persistent fibrosis in aging by targeting Nox4-Nrf2 redox imbalance. *Sci Transl Med* 2014; 6: 231ra247.
31 Hecker L, Vittal R, Jones T, *et al.* NADPH oxidase-4 mediates myofibroblast activation and fibrogenic responses to lung injury. *Nat Med* 2009; 15: 1077–1081.
32 Lee JU, Son JH, Shim EY, *et al.* Global DNA methylation pattern of fibroblasts in idiopathic pulmonary fibrosis. *DNA Cell Biol* 2019; 38: 905–914.
33 Rabinovich EI, Kapetanaki MG, Steinfeld I, *et al.* Global methylation patterns in idiopathic pulmonary fibrosis. *PLoS One* 2012; 7: e33770.
34 Cisneros J, Hagood J, Checa M, *et al.* Hypermethylation-mediated silencing of p14(ARF) in fibroblasts from idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 2012; 303: L295–L303.
35 Hanmandlu A, Zhu L, Mertens TC, *et al.* Transcriptomic and epigenetic profiling of fibroblasts in idiopathic pulmonary fibrosis (IPF). *Am J Respir Cell Mol Biol* 2022; 66: 53–63.
36 Lambert M, Jambon S, Depauw S, *et al.* Targeting transcription factors for cancer treatment. *Molecules* 2018; 23: 1479.
37 Bridges RS, Kass D, Loh K, *et al.* Gene expression profiling of pulmonary fibrosis identifies Twist1 as an antiapoptotic molecular 'rectifier' of growth factor signaling. *Am J Pathol* 2009; 175: 2351–2361.
38 Yeo SY, Lee KW, Shin D, *et al.* A positive feedback loop bi-stably activates fibroblasts. *Nat Commun* 2018; 9: 3016.
39 Palumbo-Zerr K, Soare A, Zerr P, *et al.* Composition of TWIST1 dimers regulates fibroblast activation and tissue fibrosis. *Ann Rheum Dis* 2017; 76: 244–251.
40 Plantier L, Crestani B, Wert SE, *et al.* Ectopic respiratory epithelial cell differentiation in bronchiolised distal airspaces in idiopathic pulmonary fibrosis. *Thorax* 2011; 66: 651–657.
41 Rangarajan S, Bone NB, Zmijewska AA, *et al.* Metformin reverses established lung fibrosis in a bleomycin model. *Nat Med* 2018; 24: 1121–1127.
42 Tan Q, Link PA, Meridew JA, *et al.* Spontaneous lung fibrosis resolution reveals novel antifibrotic regulators. *Am J Respir Cell Mol Biol* 2021; 64: 453–464.
43 Brooks PJ, Glogauer M, McCulloch CA. An overview of the derivation and function of multinucleated giant cells and their role in pathologic processes. *Am J Pathol* 2019; 189: 1145–1158.
44 Besnard V, Dagher R, Madjer T, *et al.* Identification of periplakin as a major regulator of lung injury and repair in mice. *JCI Insight* 2018; 3: e90163.