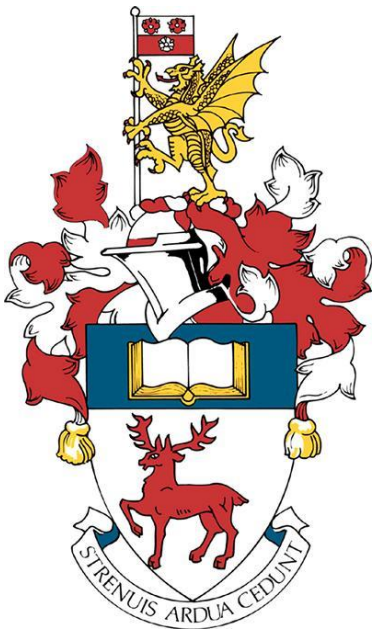
Expression and function of Wnt-inducible signalling protein 1 in idiopathic pulmonary fibrosis



University of Southampton

Faculty of Medicine

Clinical and Experimental Sciences

Joseph Alan Bell

Thesis for the Degree of Doctor of Philosophy

December 2020

Abstract

Abstract

Idiopathic pulmonary fibrosis (IPF) is a progressive, fatal fibrotic interstitial lung disease affecting older adults. It has poor prognosis and clinical interventions are limited. IPF is characterised by the build-up of fibrotic extracellular matrix in the lung interstitium, leading to progressively impaired gas exchange and eventual death. The cause of IPF is unknown but is thought to involve repeated microinjuries to lung tissue leading to chronic activation of wound healing pathways such as Wnt, TGF-β and hypoxia inducible factor (HIF) signalling. IPF tissue is characterised by the presence of fibroblastic foci – aggregates of matrix-secreting fibroblasts and myofibroblasts which drive fibrosis by producing stiffened collagen which exhibits abnormal, bone-type crosslinking.

CCN proteins are a family of proteins with similar protein structure, which affect cell signalling by interacting with other proteins to modulate their function. Wnt-inducible signalling protein 1 (WISP-1) is a Wnt-driven CCN protein that has roles in bone development and stem cell survival. While WISP-1 expression is upregulated in IPF tissue, its role in disease pathogenesis is unclear. It is hypothesised that WISP-1 is dysregulated in fibroblastic foci, and that this contributes to IPF pathogenesis by affecting protein-protein interactions. Thus, the aims of this project were to (i) characterise the localisation of WISP-1; (ii) to identify drivers of WISP-1 expression; and (iii) to identify potential functions of WISP-1 by determining what proteins it interacts with.

WISP-1 expression was found to localise to fibroblastic foci in a laser-capture microdissection (LCMD) RNAseq dataset of control and IPF alveolar septae and fibroblastic foci. This was confirmed by RNAscope *in-situ* hybridisation. WISP1 was identified as being expressed in two distinct fibroblast types in an IPF single cell RNAseq dataset, myofibroblasts and senescent, profibrotic HAS1-high fibroblasts. Gene signatures from these cells were identified in LCMD fibroblastic focus RNAseq data using CibersortX *in-silico* cell sorting. Gene expression signatures associated with HIF and TGF-β signalling were found to be upregulated in WISP1 expressing mesenchymal cells. Treatment of primary lung fibroblasts from healthy or IPF donors with HIF activators or hypoxic conditions *in vitro* led to induction of WISP1 which was greatly augmented in IPF lung fibroblasts, suggesting WISP-1 expression is HIF driven in IPF. Using an affinity purification-mass spectrometry (AP-MS) workflow to identify protein interaction partners of GFP-tagged WISP-1, the mitogen fibroblast growth factor 2 (FGF2) and the mitochondrial cell survival factors voltage dependent anion channel 1 (VDAC1) and prohibitin (PHB) were identified as WISP-1 interaction partners in MRC-5 cultured lung fibroblasts. Interactions of WISP-1 with FGF2 and VDAC1 were confirmed by western blotting or using binding assays.

Abstract

In conclusion, WISP-1 is localised in fibroblastic foci, where its expression occurs in two distinct, profibrotic fibroblast types: myofibroblasts and IPF-specific HAS-1 high fibroblasts. In vitro studies demonstrate that WISP-1 expression in IPF fibroblasts is driven by HIF signalling and that WISP-1 interacts with proteins known to affect cell survival and proliferation. These findings are consistent with a role for WISP-1 in IPF pathogenesis.

Acknowledgements

I would like to thank my supervisors for all the help and support they have offered over the course of this PhD. First, Rob Ewing, for all the Bioinformatics troubleshooting, and for helping me believe that a Biochemist could learn to code. Secondly, Jane Collins, whose invaluable advice, both scientific and otherwise, and for teaching me that the secret to science is to question everything. And finally, Donna Davies, whose expertise and tireless support has made me be the best version of myself I could be. I have yet to find a gene name she didn't know or couldn't guess. I can confidently say that without all three of them, my PhD would have been a far less rewarding experience.

I'd like to thank all the Brooke Lab members, past and present, for taking me under your wing. They were always there with useful advice, congratulations or commiserations. I am looking forward to working with them all in the future.

I'd like to thank Jay and Chiara, my fellow PhD students. We've been through a lot together, and I will always count them among my closest friends. Their humour helped me no end when I thought things were going disastrously wrong (about once a week), and they've helped make the past four years a time I will look back on with great fondness.

Thanks as well to everyone in the Ewing lab, they were always pleased to see me, even when I was parachuting in at the last minute with some inane methodology question. It was a pleasure to work with them.

Thanks to Mark Jones. His formidable expertise has helped me develop as a scientist, and shown me what it takes to be at the top of his field. His support and advice has been much appreciated throughout this project.

Thanks to all my other Southampton friends, as well as the Oxford lot, who have provided support, friendship and occasionally a welcome distraction from my PhD.

Thanks to my family, they have helped keep me sane throughout the last four years, especially during lockdown.

And finally, to Mary, thanks for everything. Her love and support has counted for more than anything else over the past four years. She has kept me going through hard times, shared in good times, and has provided a steady hand and guiding light for me since well before I started my PhD. This PhD would not have been possible without her.

Declaration of Authorship

I, Joseph Alan Bell…….

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Expression and function of Wnt-inducible signalling protein 1 in idiopathic pulmonary fibrosis

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Either none of this work has been published before submission, or parts of this work have been published as: [please list references below]:

Signed: ……

Date: ……

Table of Contents

Table of Contents

Table of Contents

List of tables

List of tables

List of figures

List of figures

Table of Abbreviations

ANOVA	Analysis of variance
AP-MS	Affinity purification-mass spectrometry
APSN	Asporin
ATI	Alveolar type I
ATII	Alveolar type II
BCA	Bicinchonic acid
BLAST	Basic local alignment search tool
CCN	Cyr61, CTGF, Nov
CDH1	E-cadherin
cDNA	Complementary DNA
CMV	Cytomegalovirus
CoIP	Coimmunoprecipitation
COMP	Cartilage oligomeric matrix protein
COPD	Chronic obstructive pulmonary disease
CPU	Central processing unit
CT	Cystine knot
CTGF	Connective tissue growth factor
CXCL	C-X-C motif ligand
DCD	Dermicidin
DKK	Dickkopf
DMEM	Dulbecco's modified Eagle's medium
DNase	Deoxyribonuclease
DTT	Dithiothreitol
ECM	Extracellular matrix
eGFP	enhanced green fluorescent protein

Table of Abbreviations

ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial-mesenchymal transition
ES	Enrichment score
FBS	Foetal bovine serum
FF	Fibroblastic foci
FGFR	Fibroblast growth factor receptor
FIH	Factor inhibiting HIF
FMT	Fibroblast to myofibroblast transition
FPKM	Fragments per kilobase per million mapped reads
FZD	Frizzled
GEO	Gene expression omnibus
GFP	Green fluorescent protein
GOI	Gene of interest
GOrilla	Gene Ontology enRIchment anaLysis and visuaLizAtion tool
GSEA	Gene set enrichment analysis
GSVA	Gene set variance analysis
HAS1	Hyaluronan synthase 1
HEK293T	Human embryonic kidney 293 cells containing SV40 T large antigen
HIF	Hypoxia inducible factor
HKG	Housekeeping gene
HRCT	High resolution computerised tomography
IGFBP	Insulin-like growth factor binding protein
IL1β	Interleukin 1 beta
ILD	Interstitial lung disease
IPF	Idiopathic pulmonary fibrosis
KEGG	Kyoto encyclopedia of genes and genomes
KRT	Keratin
LAP	Latency associated peptide
LB	Luria-Bertani
LCMD	Laser-capture microdissection
logFC	Log (base 2) fold change
LOXL2	Lysyl oxidase-like 2

Table of Abbreviations

LTBP	Latent TGF-β binding protein
MEM	Modified Eagle's medium
MMP	Matrix metalloproteinase
MRC-5	Medical research council cell strain 5
MsigDB	Molecular signatures database
MUC5B	Mucin 5B
NCBI	National centre for biotechnology information
NSIP	Non-specific interstitial pneumonitis
PCA	Principal components analysis
PCR	Polymerase chain reaction
PDGFR	Platelet-derived growth factor receptor
PGE2	Prostaglandin E2
PHB	Prohibitin
PHD	(HIF) prolyl hydroxylase
PLIN2	Perilipin 2
PLOD2	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2
POSTN	Periostin
PPARγ	Peroxisome proliferator-activated receptor gamma
RAM	Random access memory
RNAseq	RNA sequencing
ROR2	Receptor tyrosine kinase orphan receptor 2
RT	Reverse transcription
RT-qPCR	Real time quantitative polymerase chain reaction
RUNX2	Runt-related transcription factor 2
SASP	Senescence associated secretory phenotype
SERPIN	Serine protease inhibitors
SFRP	Secreted frizzled-related protein
SP	Signal peptide
SPARC	Secreted protein acidic and rich in cysteine (Osteonectin)
TBST	Tris buffered saline, 1% Tween-20
TERC	Telomerase RNA template
TERT	Telomerase reverse transcriptase
TGF-β	Transforming growth factor beta

Table of Abbreviations

TMM	Trimmed mean of M values
TNF-α	Tumour necrosis factor alpha
t-SNE	t-stochastic nearest neighbour embedding
TSP	Thrombospondin type 1 repeat
UIP	Usual interstitial pneumonia
UMAP	Uniform manifold approximation and projection
VDAC1	Voltage dependent anion channel 1
VEGFR	Vascular endothelial growth factor receptor
VWC	Von Willebrand factor type C
WISP-1	Wnt inducible signalling protein 1
Wnt	Wingless/Int-1
ZEB	Zinc finger E-box binding
α-SMA	Alpha smooth muscle actin

1 Introduction

1.1 The lung

The mammalian lung is a complex and fractal structure, evolved for efficient gas exchange. The lungs consist of airways - progressively thinner gas conducting tubes (bronchioles) branching off the bronchi, which in turn branch off the trachea, and alveoli, thin air sacs surrounded by capillaries which facilitate gas exchange between the blood and the air.

Alveoli are made up of two primary epithelial cell types, Alveolar type I (ATI) and type II (ATII) cells. ATI cells are very thin and make up the majority (96%) of the alveolar surface area. They are the principal cell type involved in gas exchange between capillaries and the airways. ATII cells are more columnar, and more numerous than ATI cells (60% of alveolar cells), despite only taking up 4% of the lung surface area(1). ATII cells are multifunctional. Firstly, they secrete surfactant, required to reduce surface tension at the air- liquid interface inside the alveoli, which helps prevent the alveoli from collapsing on themselves, and also functions in the innate immune system. Secondly, they metabolise and take up foreign substances(2). ATII cells also act as stem cells helping to repair the lung after alveolar epithelial damage, dividing and differentiating to ATI cells(1,3).

The lung interstitium, which is the area affected by idiopathic pulmonary fibrosis (IPF) and other interstitial lung diseases, is the space between the alveolar epithelium and the endothelium of the blood vessels. In healthy lungs, it is very thin, allowing efficient gas exchange between the blood and the air spaces within alveoli. The interstitium contains a network of extracellular matrix (ECM) which connects cells and tissue together within the lung interstitium(4).

1.2 Idiopathic Pulmonary Fibrosis

Idiopathic pulmonary fibrosis (IPF) is a chronic lung disease in which the interstitium of the lungs undergoes progressive fibrosis, leading to a characteristic pattern of usual interstitial pneumonia. It is one of several fibrosing interstitial lung diseases and is one of the most common interstitial lung diseases, with an estimated prevalence of up to 42.7 cases per 100,000 people in the United States(5). IPF almost exclusively affects older adults, with prevalence of IPF rising from 4 per 100,000 in the 18 – 34 year old cohort to 227.2 per 100,000 in those aged 75 years or older(5,6).

IPF diagnosis initially relies on lung function testing, as well as characteristic crackling on listening to a patient’s breathing via a stethoscope. Family history of IPF is taken into consideration, as are other potential causes of reduced lung function, such as chronic hypersensitivity pneumonitis. If

IPF is suspected, a high resolution computerised tomography (HRCT) scan can reveal the presence of honeycombing in the lungs –cystic airspaces surrounded by fibrotic tissue, cause by dilation and thickening of terminal bronchioles(7). This is characteristic of usual interstitial pneumonia (UIP) – a heterogeneous phenotype where areas of dense fibrosis at the lung periphery give way to a more normal lung appearance(8). The microscopic network of alveolar epithelium which allows efficient gas exchange with red blood cells is thickened, leading to impaired gas exchange in the lungs of IPF sufferers(9). Recognition of UIP is important for diagnosis of IPF, but other interstitial lung diseases may also display features of UIP. To positively diagnose IPF, a surgical biopsy of lung tissue is sometimes required. Biopsied tissue is stained with haematoxylin and eosin, which stain for nuclear matter and extracellular matrix respectively(8,9).

UIP is characterised by the presence of fibroblastic foci – aggregates of fibroblasts which produce collagenous extracellular matrix (ECM). The presence of fibroblastic foci is an important diagnostic criterion for IPF. Fibroblastic foci contribute to disease progression by increasing ECM production(9,10). Figure 1.1 shows a lung HRCT scan with prominent honeycombing, and a micrograph of IPF lung showing characteristics of UIP, such as areas of fibrotic tissue and a fibroblastic focus.

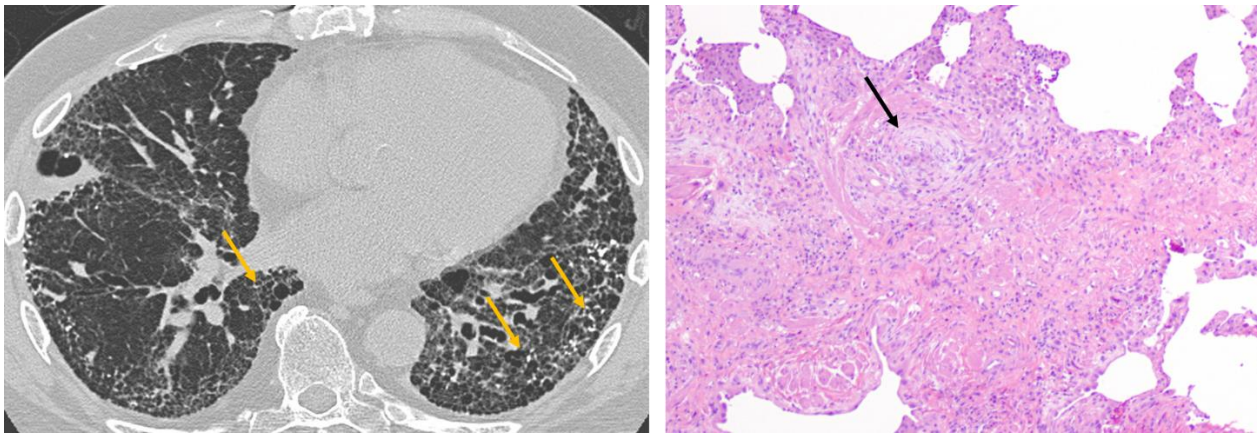


Figure 1.1: Diagnosis of IPF. Left: HRCT scan of IPF lung showing lung honeycombing (yellow arrows). Image taken by Yale Rosen, reproduced here under Creative Commons CC BY-SA 2.0 license(11). Right: haematoxylin/eosin stain of IPF tissue, showing usual interstitial pneumonia Blue stain is haematoxylin, showing cell nuclei, pink is eosin, staining extracellular matrix. Arrow shows a fibroblastic focus. Image is public domain, taken from https://en.wikipedia.org/wiki/Usual_interstitial_pneumonia (11)

IPF has poor prognosis, with median survival from diagnosis being three years. Accumulation of scar tissue leads to a decrease in respiratory capacity, shortness of breath and eventual respiratory failure. Disease progression is irreversible(6).

Disease progression of IPF is not linear; many patients experience short-term rapid worsening of lung function known as acute exacerbations which contrast sharply with the normal slow disease progression(6). The cause of acute exacerbations is unknown, but they may correlate with respiratory viral infection or mechanical stress in the lungs leading to pathogenic activation of wound-healing pathways(12).

1.3 Environmental risk factors for IPF

The cause of IPF is unknown, but several environmental risk factors are associated with IPF. A 1997 case-control study demonstrated a link between cigarette smoking and IPF. Individuals with a history of smoking are more likely to develop IPF(13). Several other studies have demonstrated less conclusive associations with environmental insults and IPF, including dust exposure from stone, wood or metalworking, and exposure to livestock and other agriculture(14). The diverse range of environmental exposures associated with IPF risk suggests that IPF is a heterogeneous disease where a large number of different sources of lung damage can lead to development of similar symptoms(14).

1.4 Genetic risk factors for IPF

Several previous studies have demonstrated genetic factors that increase the risk of IPF. A variant of the MUC5B gene, which codes for a mucin molecule which is a component of airway mucus, has been identified as being more common in IPF(15). MUC5B has also been implicated in host-defence of the airways, where it may play a role in infection control(16). Other loci identified in large-scale genome-wide association studies are the TERT and TERC genes, which encode part of the telomerase enzymes associated with cellular renewal and loss of senescence, and the desmoplakin (DSP) gene, which codes for a component of desmosomes which contributes to epithelial barrier integrity(15,17).

1.5 Treatments for IPF

IPF has no cure. However, two treatments which slow disease progression, pirfenidone and nintedanib, have entered the clinic(18). Pirfenidone is an anti-fibrotic drug which reduces fibroblast proliferation and activation, leading to decreased fibrosis(19,20). Nintedanib is an inhibitor of several growth factor receptors which contribute to the proliferation of fibroblasts in IPF, including platelet-derived growth factor receptor (PDGFR), vascular endothelial growth factor receptor (VEGFR) and several fibroblast growth factor receptors (FGFR 1-4)(21).

Pirfenidone was first identified as a compound with potential antifibrotic properties in a bleomycin-induced hamster model of IPF(19,22). It displays antifibrotic properties in a broad array of animal models, and in clinical trials it was well tolerated by IPF patients and reduced the number of acute exacerbations in pirfenidone treated patients(18). Its mechanism of action is obscure, although pirfenidone treatment does reduce the expression of several known profibrotic molecules in animal models of IPF(19). In the bleomycin mouse model, it reduces expression of TGF-β, the prototypical profibrotic cytokine, as well as several matrix metalloproteinase (MMP) genes, and profibrotic cytokines and chemokines like CXCL12(23). Cell-based assays of pirfenidone's effect demonstrated a reduction in profibrotic responses in cultured dermal myofibroblasts(24).

Nintedanib has also been shown to have a broad range of antifibrotic effects(18). However, in contrast to pirfenidone, nintedanib's mechanism of action is known(21). It is a small-molecule tyrosine kinase inhibitor of growth factor receptors, particularly vascular endothelial growth factor (VEGF) receptor and fibroblast growth factor (FGF) receptors(21).

1.6 Animal models of IPF

The principal animal model used for studying IPF is the bleomycin-induced mouse model. Bleomycin is a chemotherapeutic agent that rapidly induces fibrosis in the lungs of treated mice through the production of reactive oxygen species, causing DNA damage. A single, intratracheal bleomycin dose in mice leads to an inflammatory response within the first week which then develops into fibrosis over a two-week period(25). As a model, the bleomycin mouse has seen notable success, including identifying TGF-β as a major driver of fibrosis(26). However, it also has several shortcomings, including the induction of an inflammatory response not seen in IPF in humans, and the bleomycin-induced fibrosis is reversible whereas the human disease is progressive and irreversible(25). While there has been considerable success in identifying many anti-fibrotic agents in this model, with the exception of nintedanib, this success has not translated into clinically relevant treatments(27,28).

Other mouse models of fibrosis exist, including exposure of mice to silica. Mouse lungs which have been exposed to mineral fibres exhibit hallmarks of fibrosis, including development of fibrotic lesions similar to those seen in human lungs exposed to silica. This model also shows upregulation of profibrotic cytokines, including TGF-β, and lesions persist for long periods of time, allowing long-term studies of disease progression(28,29). However, these mice do not develop fibroblastic foci, so the model has limited utility for IPF. As IPF is disease of aging, some mouse models have used aged mice to try and mimic the effects of pulmonary fibrosis – one study

identified that pulmonary fibrosis develops in response to γ-herpesvirus exposure in aged mice. This may reflect differences in lung tissue response to viral infection due to fibrosis(30). Treatment of mice with expression vectors containing the genes for specific cytokines can allow investigation of the effect of that specific cytokine in driving fibrosis – this has had success in showing that prolonged TGF-β expression in mouse lungs can drive fibrosis, for example(31). However, single-cytokine models are necessarily reductive, and do not recapitulate the complexity of human IPF. Other animals also develop fibrosing interstitial lung diseases, but they do not closely resemble IPF(28).

1.7 Mechanisms of Pathogenesis

IPF is a complex disease with unknown cause, but several mechanisms have been described as contributing to pathogenesis. Firstly, injuries to the alveolar epithelium lead to activation of repair pathways, such as Wnt and TGF-β signalling(32,33). In a normal wound healing response, this activation would end after tissue repair was complete, but in IPF these pathways continue signalling, leading to pathogenesis(32). It is likely that this continued signalling by repair pathways is the result of multiple tissue microinjuries. This chronic, deregulated activation of repair pathways is known to contribute to disease progression(33,34). TGF-β signalling has been implicated in epithelial to mesenchymal transition of alveolar epithelial cells, fibroblast activation and differentiation to myofibroblasts, and reduced proliferation and apoptosis of epithelial cells, leading to a reduction in re-epithelialisation of the lung after tissue injury(10,34).

The alveolar epithelium has been implicated in pathogenesis, with alveolar epithelial cells being the source of TGF-β, Wnt ligands and chemotactic factors such as CXCL12 which recruit fibrocytes to the lung(35,36). Alveolar cells also produce the α5β6 integrins; these activate latent TGF-β by associating with the latency associated peptide (LAP) of TGF-β, leading to cleavage and TGF-β activation(37,38). Wound healing in non-diseased lung tissue requires efficient re-epithelialisation of injured alveolar epithelium(39). AECs in IPF are less efficient at this re-epithelialisation, possibly due to increased apoptosis of AECs and reduction in the ability of ATII cells to differentiate to ATI cells, the cell type which mediates gas exchange in the lung, as well as loss of basement membrane integrity(40).

1.8 Fibroblasts in IPF

Fibroblasts are an important non-immune cell found in the lung interstitium. Fibroblasts are important for production and regulation of the lung extracellular matrix (ECM)(41). ECM in the lungs is important for providing stability and elastic recoil, both of which are crucial for the

respiration cycle. ECM in the lungs is divided into basement membranes – thin, dense sheets of matrix which epithelial cells adhere to – and interstitial ECM, which is a less dense network connecting the components of the lung interstitium together(42). In the healthy lung, this network allows significant expansion and contraction of the lung associated with normal breathing. ECM proteins include collagens, which provide most of the structure of the lung interstitium, as well as elastins which form elastic fibres, allowing the interstitium to stretch and contract during breathing(42). ECM is maintained by resident pulmonary fibroblasts. It is heavily modified in IPF, with thickened, stiffened fibrotic epithelium preventing elastic recoil, and impairing gas exchange(42).

1.9 Fibroblastic foci in IPF

Fibroblasts are crucial for maintaining lung homeostasis in healthy lung tissue, by maintaining the interstitial ECM. However, in idiopathic pulmonary fibrosis, they are the cells which secrete the thickened, highly fibrotic ECM which is characteristic of the disease. Fibroblasts in the IPF lung aggregate to form fibroblastic foci, areas of proliferating, highly secretory fibroblasts and myofibroblasts. These are a key histological marker of IPF(43). Fibroblastic foci are discrete aggregates of proliferating fibroblasts – a micro computerised tomography (micro-CT) study of IPF lung tissue showed that fibroblastic foci do not form an interconnected reticulum, but were in fact separate cellular aggregates embedded in the lung interstitium(43). This contradicted previous findings modelling focus interconnectedness using highly z-stacked images(44,45).

1.10 Source of fibroblasts in IPF

Although the source of the fibroblasts which make up fibroblast foci is not known with certainty, and is likely a combination of different factors, three potential sources are epithelial to mesenchymal transition (EMT) of alveolar epithelial cell, migration and differentiation of bone marrow fibrocytes from the blood to fibroblastic foci, and differentiation of mesenchymal progenitor cells already present within the lung(46–49).

1.11 Myofibroblasts in IPF

Myofibroblasts are a subset of fibroblasts characterised by their expression of alpha smooth muscle actin (α-SMA). This is an isoform of the contractile muscle protein actin which is expressed in vascular smooth muscle cells, allowing contraction and expansion of vascular tissue. In myofibroblasts, this α-SMA forms stress fibres(47). Myofibroblasts are thought to differentiate from existing tissue-resident fibroblasts during normal wound-healing processes, where they

produce large amounts of extracellular matrix to facilitate tissue repair, and their contractile properties help close wounds(47). In normal wound healing, the presence of myofibroblasts is temporary, whereas in fibrosis they are more permanent and resistant to apoptosis(50). This resistance to programmed cell death leads to the accumulation of myofibroblasts in the fibrotic lung, and the eventual formation of fibroblastic foci(50,51).

The principal driver for the myofibroblast activation and proliferation seen in both normal wound-healing responses and fibrosis is transforming growth factor beta (TGF-β). TGF-β drives myofibroblast transdifferentiation, stimulating α-SMA production and the formation of contractile stress fibres(52). The contraction of these filaments exerts a mechanical stretching force on nearby cells via integrin-containing focal adhesion complexes. This mechanical stretch can activate latent TGF-β associated with the extracellular matrix, in a positive feedback loop which continually sustains the myofibroblast phenotype(53).

1.12 Other fibroblast types

New technologies such as RNAseq analysis have enabled identification of other fibroblast populations in the lung. In particular, rodent lungs contain lipofibroblasts – quiescent fibroblasts which are thought to contribute to the maintenance of the integrity of the epithelial barrier and are important for lung homeostasis and repair(54). Lipofibroblasts are characterised by the presence of lipid droplets within the cells, and they are most abundant in developing rodent lungs(55). Their roles in the lung include surfactant synthesis, crucial for epithelial barrier integrity. Lipofibroblasts store triglycerides in their lipid droplets and release them in response to prostaglandin E2 (PGE2) release from ATII cells. This supports synthesis of surfactant, helping maintain epithelial barrier integrity and supporting lung function(54,56). They are also key sites of retinoid accumulation during lung development, and they produce various key ECM proteins, including elastins and type I collagen. Their presence in mice and rats is well-attested, however their existence in humans is controversial(57).

A recent single-cell RNA-seq experiment which extracted and sequenced cells from lung tissue derived from both control lung and a variety of different interstitial lung diseases, including IPF, identified several different fibroblast subtypes which may be of significance both to normal lung function and IPF pathogenesis(58). As well as identifying myofibroblasts and a more normal subset of fibroblasts, likely corresponding to undifferentiated interstitial fibroblasts, this study also identified a subset of fibroblasts which highly express the perilipin 2 (*PLIN2*) gene. Perilipin 2 is an important protein associated with lipid packaging, and the paper suggests that these cells may be comparable in function to mouse lipofibroblasts. Another population identified in this

study was a subset of fibroblasts enriched for hyaluronan synthase 1 (*HAS1*). These cells were enriched for pathways associated with cellular stress as well as the epithelial to mesenchymal transition markers SNAI1 and TWIST. Immunofluorescent staining revealed HAS1-high cells localise further away from the pleural surface than other fibroblast types, suggesting that they may be evidence of a migratory fibroblast phenotype from other tissue areas(58).

Figure 1.2 shows a schematic of a normal and IPF alveolus, showing the major features of normal and IPF lung tissue.

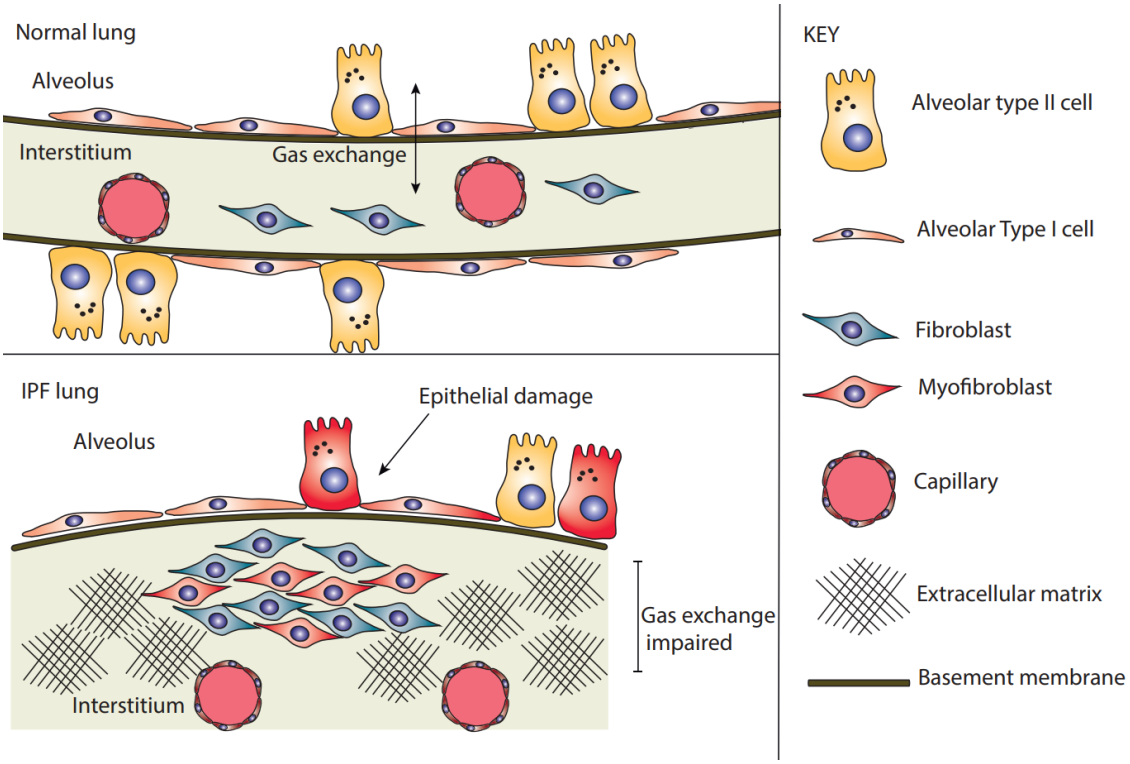


Figure 1.2: Cross section of a normal and IPF alveolus. Top: Normal alveolar epithelium, allowing efficient gas exchange. Bottom: IPF alveolar epithelium, showing fibroblastic foci underlying abnormal epithelium, and thickened, stiffened matrix impairing gas exchange.

1.13 Extracellular matrix in normal lung interstitium

The extracellular matrix in the normal lung interstitium is composed mostly of collagen(59,60). Collagens are a diverse range of structural proteins which spontaneously assemble into triple helices consisting of three collagen monomers. 30 different types of collagen have been described in humans, with 11 of these being fibril forming, or fibrillar collagens(61). These assemble into long fibrils, whose high tensile strength lends structure to the lung interstitium(61). The stiffness of these collagen fibrils is largely determined by the amount and type of cross linking between

collagen molecules. Collagen crosslinking occurs on the collagen telopeptide, the non-helical portion of the molecule, between lysine residues on adjacent telopeptides. Two distinct forms of cross-linking occur, based on whether or not lysine residues are converted to hydroxylysine. The conversion of lysine to hydroxylysine residues is catalysed by lysyl hydroxylase enzymes (procollagen-lysine-5-dioxygenases, PLODs). Non-hydroxylated lysine residues are converted to lysine aldehydes by lysyl oxidase (LOX) enzymes. These lysine aldehydes spontaneously react with lysine or hydroxylysine to form aldimines, which then react with histidine residues in helical collagen to form histidinohydroxylysinonorleucine (HHL) crosslinks. These crosslinks are more prevalent in collagen in the skin and tendons. LOX enzymes convert hydroxylated lysine residues to aldehydes as well, but these react with lysine or hydroxylysine to form a keto-imine. This keto-imine then reacts further to form pyridinoline or deoxypyridinoline crosslinks. These crosslinks are characteristic of the collagen found in bone(62–64). Abnormal collagen crosslinking characterised by an increase in these bone-type pyridinoline crosslinks is thought to be a major contributor to the increased stiffness of the extracellular matrix seen in IPF. Jones et al. showed that tissue stiffness was increased in IPF, independent of the amount of collagen present. This study also demonstrated increased expression of the lysyl oxidase enzyme genes LOXL2, LOXL3 and LOXL4, as well as lysyl hydroxylase 2 (PLOD2), and increased hydroxylysine bone-type collagen crosslinking, which correlated well with tissue stiffness(65). The lung interstitium contains mostly fibrillar collagen types I and III, with these collagens defining alveolar structure, and thus the microarchitecture of lung tissue(42,66).

Elastic fibres are another important constituent of the lung ECM, contributing elasticity to lung tissue, allowing the stretching and contraction of ECM required for breathing. These consist of the ECM proteins elastin, fibrillin and fibulin(42,67).

The final principal component of lung extracellular matrix is a multitude of proteoglycans. These form a hydrated gel-like structure in which the structural proteins described above are embedded. Important components of this structure include glycosaminoglycans (GAGs), chondroitin sulphate and heparan sulphate(68).

1.14 Extracellular matrix in IPF

The extracellular matrix found in IPF tissue differs from the that found in normal, healthy lung in several ways. Firstly, it is stiffer, a feature that has been linked to an increase in bone-type, pyridinoline collagen crosslinks(65,69). This lung stiffening leads to shortness of breath in IPF patients, often the first clinical indication of an eventual IPF diagnosis(10). Interestingly, Jones et al. found an increase in collagen stiffness in IPF, but not in the amount of collagen compared to

control lung, suggesting that it is primarily collagen stiffness that contributes to IPF pathogenesis(65).

This study identified an increase in the activity of amine oxidase collagen crosslinking enzymes, and that increased enzymatic activity correlated with increased lung tissue stiffness. They showed that selective inhibition of the collagen crosslinking enzymes LOXL2 and LOXL3 reduces pyridinoline crosslinking density and tissue stiffness in a 3D cell culture model of IPF fibroblastic foci, and demonstrated reversal of fibrosis and return to normal lung structure in a rat model of pulmonary fibrosis(65).

A gene expression profiling study of the extracellular matrix in IPF did show an increase in collagen gene expression, however, as well as an increase in soluble factors which may contribute to extracellular matrix remodelling, notably the matrix metalloprotease enzymes (MMPs)(70). Humans have 24 MMP genes, which produce secreted enzymes with roles in extracellular matrix remodelling and ECM protein degradation. They can also play a role in signalling, cleaving proteins which regulate ECM production such as osteopontin(71). In IPF, MMP7 expression is upregulated, and the level of MM7 expression is prognostic for how fast the disease will progress(72,73), while MMP7-deficient mice are resistant to bleomycin-induced fibrosis(74). Proposed profibrotic roles for MMP7 include cleavage of basement membrane proteins, leading to degradation of lung architecture, as well as cleavage of substrates such as Fas ligand and E-cadherin leading to profibrotic signalling(75,76). Other MMPs, such as MMP8, may contribute to fibrocyte migration into lung tissue, leading to increased fibroblast numbers in IPF, while still others, such as MMP19, may have a protective, antifibrotic effect(76).

A transcriptional study that used laser capture microdissection to excise fibroblastic foci in IPF, followed by RNAseq analysis to identify gene expression signatures in excised tissue found a collagen-associated gene signature in fibroblastic foci. This both confirms that fibroblastic foci are the sites of collagen gene expression, as well as defining several gene sets associated with collagen expression, including TGF-β associated genes(77). A study looking at collagen staining of lung tissue from lung biopsies also identified staining for the collagen C-terminal telopeptide around fibroblastic foci, again identifying fibroblastic foci as areas of collagen deposition(78). Fibroblastic foci are the centres of collagen deposition, and extracellular matrix staining of fibroblastic foci reveals intense staining within and around them(79).

1.15 Mechanical stretch of lung tissue in IPF induces TGF-β release

Recent research has highlighted the role of stiffened, fibrotic tissue in IPF. Lung tissue strips subjected to mechanical stretching released TGF-β, with fibrotic lung tissue releasing more TGF-β(53). This is significant for pulmonary fibrosis given the increased tissue stiffness seen in fibrotic lungs, leading to an increase in the effective stretching force lung cells experience during breathing. Mechanical stretch can mediate profibrotic effects via yes-associated protein/tafazzin (YAP/TAZ) signalling, with activation of this signalling pathway leading to increased production of profibrotic molecules, including connective tissue growth factor (CTGF) and type I collagen (COL1A1)(80). This creates a positive feedback loop, where increased myofibroblast differentiation and activation in fibroblastic foci leads to increased collagen stiffness, leading to TGF-β release and YAP/TAZ signalling, which further promotes myofibroblast activation and fibroblast activation(69,81).

1.16 TGF-β in IPF

TGF-β is an important contributor to IPF pathogenesis. It is a profibrotic cytokine which leads to fibrosis. Mechanisms by which TGF-β contributes to fibrosis include induction of epithelial to mesenchymal transition (EMT) in epithelial cells, recruitment of circulating fibroblast precursor cells from the bone marrow, and activation of already-present pulmonary fibroblasts. This causes them to differentiate, secrete more extracellular matrix and adopt a myofibroblast phenotype (fibroblast to myofibroblast transition, FMT)(82).

TGF-β has three isoforms, TGF-β1, -2 and -3. TGF-β1 is the isoform most associated with IPF. It is secreted in a latent form as a peptide chain containing the TGF-β protein and a latency associated peptide (LAP). Latent TGF-β is also bound to latent TGF-β binding protein (LTBP). Multiple mechanisms for TGF-β release from its latent form (i.e. cleavage of the LAP and removal of LTBP) have been described, including proteolytic cleavage by the matrix metalloproteases MMP9 and MMP2, tissue stiffness, thrombospondin-1 and several integrins, which bind the latency associated peptide and may aid recruitment of MMPs(83).

TGF-β is produced by several different cell types in IPF, including epithelial cells, fibroblasts and myofibroblasts and macrophages. (84). Figure 1.3 shows a schematic of how TGF-β signal transduction occurs. TGF-β binds type I and II cell surface receptors, forming a ligand-receptor

complex. Type II receptors phosphorylate type I receptors, which in turn phosphorylate receptor regulated SMAD proteins, the principal intracellular effectors of TGF-β signalling(85).

Receptor-regulated SMAD proteins (R-SMADs), consisting of the proteins SMAD1-3 SMAD5 and SMAD8/9, once phosphorylated, bind to SMAD4, with two R-SMAD and one SMAD4 molecule forming a heterotrimeric complex. This R-SMAD/SMAD4 complex enters the nucleus and is capable of regulating gene expression by binding to DNA in combination with various transcription factors and chromatin remodelling enzymes.

TGF-β signalling in IPF is associated with a number of different profibrotic processes, including increased expression of extracellular matrix genes, including multiple collagen genes, leading to extracellular matrix remodelling in IPF(86,87). TGF-β signalling is also associated with the conversion of fibroblasts to myofibroblasts and myofibroblast activation(47,52). For alveolar epithelial cells, TGF-β signalling has been shown to increase epithelial-mesenchymal transition in IPF, and has been implicated in apoptosis of alveolar epithelial cells, a key component of the alveolar destruction associated with pulmonary fibrosis(88,89).

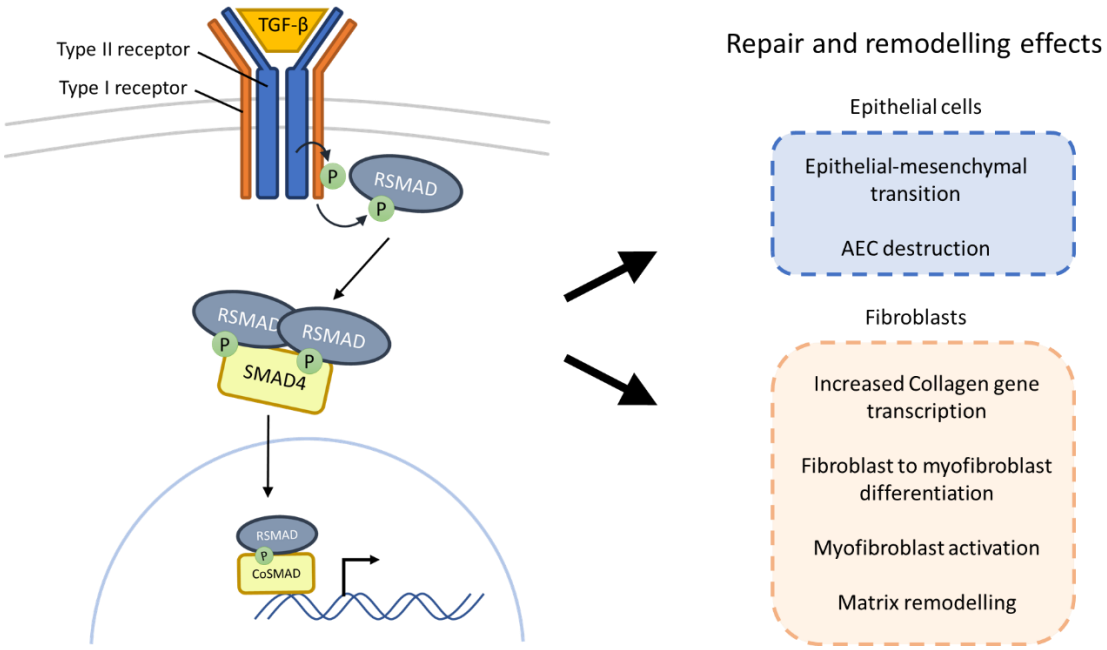


Figure 1.3: Schematic showing the TGF-β signalling pathway and its roles in repair and remodelling in lung epithelial cells and fibroblasts.

1.17 Epithelial-mesenchymal transition in IPF

Another process which has been implicated in fibrosis is epithelial to mesenchymal transition (EMT). This is the process where epithelial cells, such as ATII cells, change their phenotype, becoming more migratory, resistant to apoptosis, and producing more ECM components. It is

characterised by the loss of expression of epithelial markers, such as E-cadherin, in favour of mesenchymal markers such as vimentin and twist. EMT can be driven by several different factors, including fibroblast growth factors and TGF-β(90). EMT occurs during the normal process of epithelial repair following injury and it has been implicated in IPF pathogenesis(91). ATII cells have been shown to undergo EMT in bleomycin induced fibrosis, and EMT of ATII cells becoming fibroblastic has been proposed as a mechanism for the increased number of fibroblasts associated with IPF(92,93). However, lineage-tracing experiments could not identify differentiation of ATII cells to fibroblasts in a bleomycin-induced mouse model of fibrosis(48), suggesting that even if EMT does occur, it is not a significant contributor to fibroblast populations in IPF.

1.18 Hypoxia in IPF

Another driver of IPF is hypoxia. Although normally an oxygen-rich environment, the lungs of people with IPF become hypoxic due to the severely impaired gas exchange associated with the disease(94). This hypoxia has been shown to contribute to IPF pathogenesis, and recent research has highlighted how hypoxia may contribute to lung fibrosis in IPF(95–97).

The body's response to hypoxia is regulated by the hypoxia inducible factor (HIF) proteins, with the HIF1α and HIF2α proteins being key master regulators of HIF activation(98). These are transcription factors which undergo post-translational modifications in response to cellular oxygen levels. In particular, the prolyl hydroxylase (PHD) enzymes hydroxylate HIF-α molecules, targeting them for proteasomal degradation(99). These enzymes are α-ketoglutarate dependent dioxygenases whose activity depends on cellular O_2 levels, and thus their enzyme activity is primarily regulated by the oxygen availability within cells(100). Low cellular oxygen levels therefore lead to less hydroxylation of HIF and therefore an increase in HIF signalling. Proteasomal degradation of HIF proteins is facilitated by ubiquitination of HIF-α proteins and binding of Von Hippel-Lindau (VHL) proteins(99). The HIF1α and HIF2α proteins form dimers with HIF1β, a constitutively expressed HIF protein which is significantly less labile than the alpha subunits. These dimers enter the nucleus and induce gene expression by binding to hypoxia response elements (HREs) in the genome(99). HIF1α is expressed in all tissue types, whereas HIF2α appears to have a more limited expression distribution(101). The two proteins induce expression of overlapping, but distinct genes(102,103). Figure 1.4 shows a schematic of HIF signalling.

The downstream effects of HIF activation are varied, but common effects include upregulation of glycolysis and cellular glucose uptake via increased expression of transmembrane glucose transporters and glycolytic enzymes(103). HIF signalling can also reduce mitochondrial oxidative

phosphorylation in favour of glycolysis, by upregulating expression of pyruvate dehydrogenase kinase 1, which negatively regulates enzymatic activity of pyruvate dehydrogenase. This blocks pyruvate to acetyl-coenzyme A production, reducing Krebs cycle activity and oxidative phosphorylation, thereby reducing O_2 consumption as dioxygen is the final electron acceptor of the mitochondrial electron transport chain(103).

A 2019 study by Aquino-Galvéz et al. demonstrated that the levels of HIF subunits HIF1α and HIF2α were both upregulated in cultured IPF fibroblasts and IPF lung tissue than control lung tissue(95). Other research has identified hypoxia as being important in the differentiation of fibroblasts to myofibroblasts, with pulmonary fibroblasts grown in hypoxic conditions showing increased expression of the myofibroblast marker α-SMA, and suppression of Thy-1, a regulator of cell-cell and cell-ECM interactions(97). These changes were linked to epigenetic modifications to the fibroblast genome, specifically hypermethylation of the Thy-1 promoter region, leading to an increase in pro-fibrotic gene expression and a more myofibroblastic phenotype(97).

Hypoxia has also implicated in the increased tissue stiffness seen in IPF. Brereton et al. demonstrated that the expression of the collagen crosslinking enzymes LOX and LOXL2, which have been previously shown to be the drivers of the increased collagen stiffness seen in pulmonary fibrosis, is driven by HIF pathway activation. For this study, a prolyl hydroxylase inhibitor was used to mimic the effects of hypoxia by reducing HIF hydroxylation and degradation(96).

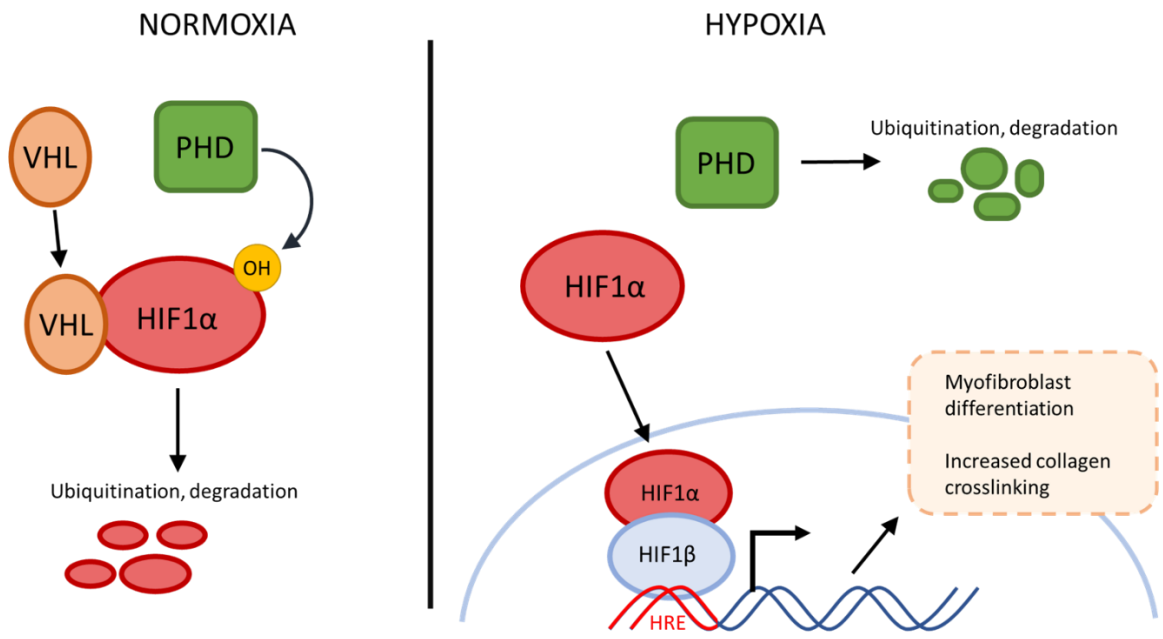


Figure 1.4: Schematic of HIF signalling pathway in normoxia (left) and hypoxia (right), with effects known to occur in pulmonary fibrosis shown.

1.19 Wnt signalling in IPF

Wnt signalling is a signalling pathway implicated in development and disease which was first discovered in mice and fruit flies(104). Wnt signalling is a highly conserved pathway common to all multicellular animals. Wnt signalling in mammals is mediated via 19 Wnt ligands – small, cysteine rich proteins homologous to the *Drosophila melanogaster* protein Wingless – secreted into the extracellular space. Wnt ligands bind to the transmembrane receptors frizzled and Low-density lipoprotein receptor-related protein 5/6 (LRP5/6). These then signal inside the cell through a variety of different pathways. Which pathway is stimulated depends on the particular Wnt ligand present, with canonical Wnt signalling being the most well-characterised(104). Figure 1.5 shows how canonical Wnt signalling works.

Non-canonical Wnt signalling involves Wnt pathways that mediate their function independently of β-catenin. The best characterised non-canonical Wnt signalling pathway is the planar cell polarity (PCP) pathway. Wnt signalling through frizzled receptors leads to activation of Ras-family GTPases. This leads to downstream changes in cytoskeletal development. The PCP pathway is important for regulating cell polarity and is important for embryonic development(105,106).

The other non-canonical Wnt signalling pathway is the Wnt/Ca^{2+} pathway. This also functions via frizzled proteins but involves activation of heterotrimeric G-proteins. These activate phospholipase C, which increases intracellular Ca^{2+} release, leading to changes in calcium regulated processes such as cell migration and cell fate determination(105,107).

Wnt signalling is an important pathway regulating wound healing and tissue regeneration in all higher organisms. Activation of Wnt signalling leads to recruitment of progenitor cells to a wound site(108). These cells can then proliferate and regenerate the damaged tissue. The initial activation of Wnt signalling in wound healing has been linked to hypoxia, and specifically HIF signalling. A study on stem cells showed increased Wnt activity in areas of hypoxic tissue, which was abrogated on HIF1α deletion, while TCF/LEF Wnt-mediated transcription and β-catenin activation was upregulated in embryonic stem cells grown in hypoxic conditions(109). As HIF signalling is present in fibrotic tissue(110), chronic Wnt activation could be maintained by HIF signalling.

Wnt signalling has been shown to be important for IPF pathogenesis. β-catenin, the principal intracellular mediator of canonical Wnt signalling, was shown to accumulate in the nuclei of cells in IPF epithelia, bronchial proliferative lesions and fibroblastic foci(111). Other evidence that Wnt signalling is important in disease is the upregulation of Wnt pathway genes in IPF in large-scale microarray studies, as well as an increase in expression of Wnt-responsive genes, including Wnt inducible signalling protein 1 (WISP-1), the principal subject of this thesis(72,112). A 2012 study indicated that canonical Wnt signalling was activated in skin fibrosis, and that this was required for TGF-β mediated fibrosis to occur(113). The same study also suggested that TGF-β was a principal activator of Wnt signalling in fibrotic diseases. Knockdown of the canonical Wnt co-receptor LRP5 was shown to be protective against bleomycin-induced pulmonary fibrosis in mice. This did not prevent against TGF-β-induced fibrosis when LGR5-deficient cells were transplanted into wild-type mice, suggesting that Wnt signalling may regulate TGF-β production in the IPF lung(114).

A quantitative real-time PCR (qPCR) study of Wnt signalling components in IPF tissue found that Wnt1, Wnt7b and Wnt10b expression were all upregulated in IPF lung(113). The frizzled Wnt receptor proteins Fzd2 and 3 were also upregulated in the IPF lung, as was β-catenin, the primary intracellular effector of canonical Wnt signalling, and the Wnt transcription factor Lef1, as well as an upregulation of Wnt target genes including MMP7 in IPF. The same study attempted to localise the Wnt signalling components using immunohistochemistry – they demonstrated an increase in Wnt signalling in alveolar epithelial type II cells (ATII), with large amounts of nuclear β-catenin staining in hyperplastic ATII cells, especially those near sites of bronchiolised tissue(113).

Wnt signalling is known to be important for wound healing and epithelial cell repair in the IPF lung. The current paradigm for the induction of IPF is microinjuries to the alveolar epithelium leading to activation of repair pathways – the induction of Wnt signalling in the IPF epithelium at

sites of alveolar hyperplasia and tissue injury suggests that Wnt is one of the principal repair pathways aberrantly activated in IPF. A study of bleomycin-induced pulmonary fibrosis in mice showed that β-catenin is a survival factor for alveolar cells in fibrosis(115). Downstream effects of Wnt signalling in IPF include induction of WISP-1 expression as well as increased interleukin-1β (IL-1β) and interleukin-6 expression(116). IL-1β has previously been shown to be a pro-fibrotic cytokine in bleomycin-induced mouse models of pulmonary fibrosis, and IL-6 has been shown to be a proliferative factor for human fibroblasts(117).

In summary, Wnt signalling in IPF is likely a consequence of tissue injury leading to activation of the Wnt repair pathway in IPF, which activated pro-fibrotic pathways leading to pulmonary fibrosis.

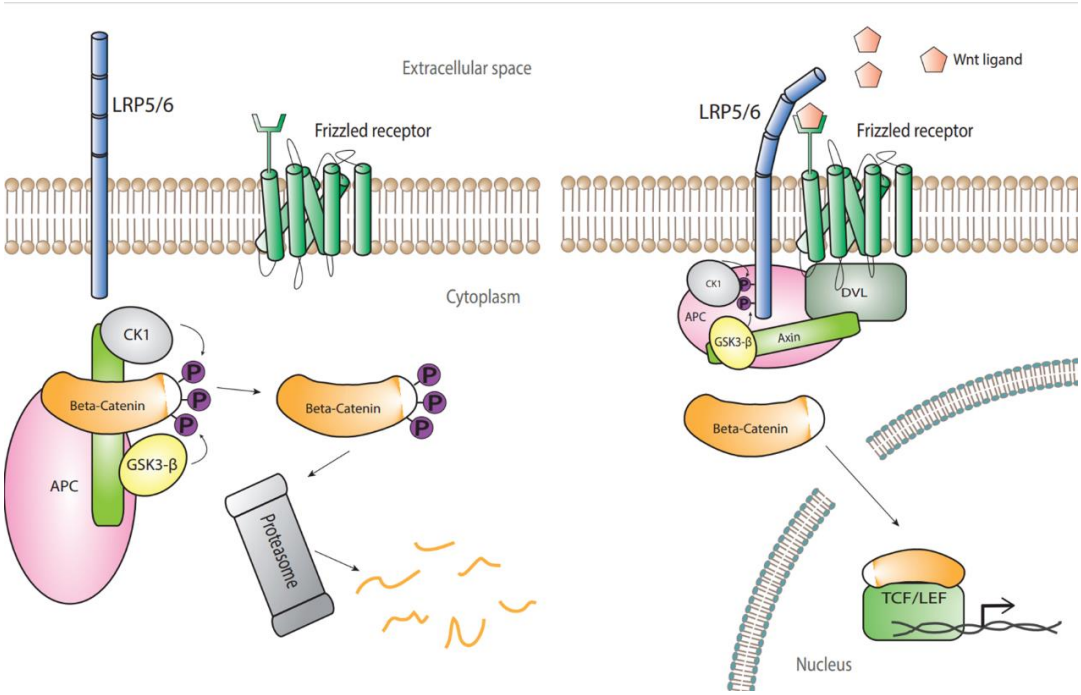


Figure 1.5: Canonical Wnt signalling. Left: If no Wnt signalling molecules are present, β catenin is targeted for degradation by the destruction complex. Right: If a Wnt ligand binds to the frizzled and LRP5/6 receptors, β catenin is phosphorylated, and can translocate into the nucleus and activate Wnt-responsive genes.

1.20 Wnt inducible signalling protein 1 (WISP-1) and the CCN proteins

Wnt inducible signalling protein 1 (WISP-1) is a matricellular protein important for bone and tissue development. Matricellular proteins are extracellular proteins which do not play a structural role

– they are important for mediating signalling between cells and their environment(118). WISP-1 is part of a larger family of six CCN proteins, named after its first three discovered members, CYR61, CTGF and NOV, which exhibit a conserved domain structure, consisting (from N-to C terminus) of an N-terminal signal peptide targeting them for extracellular secretion, an insulin-like growth factor binding domain (IGFBP), a von Willebrand Factor type C domain (VWC), a thrombospondin type I repeat (TSP) domain and a cystine knot (CT) domain towards the C-terminus(119). Figure 1.6 shows the domain structure of CCN proteins, as well as interaction partners which have been verified for at least one CCN protein. Roles of CCN proteins include regulating cell-matrix adherence, ECM remodelling, wound repair, cell proliferation and angiogenesis(119). They have also been implicated in bone development(120).

CCN proteins are known to interact with other proteins, including integrins, bone morphogenic proteins, TGF-β and multiple extracellular proteins(121,122) (Figure 1.6). Several CCN proteins interact with integrins, including Cyr61 (CCN1), which has been shown to bind to α5β6 integrins, and can mediate senescence in fibroblasts via an α6β1 integrin and heparan sulphate proteoglycan dependent interaction(123,124). Connective tissue growth factor (CTGF, CCN2) also interacts with α5β6 integrins. CTGF knockout in mice also abrogated liver fibrosis, and the inhibition of the α5β6 integrin-CTGF interaction via a neutralising antibody led to a reduction in expression of profibrotic mediators, including α-SMA, TGF-β and procollagen-α1(125). CTGF has also been shown to be important for fibrosis, and it has been identified as a downstream mediator of TGF-β-induced myofibroblast activation(126,127). While CCN proteins are classically known as secreted, matricellular proteins, they also have been shown to have intracellular signalling roles. CTGF has been shown to signal intracellularly after uptake by human mesangial cells, (128). The human protein atlas also characterises WISP1 as localising both to the cytosol and the extracellular space within tissue(129).

CCN proteins exhibit similarities at both the domain and sequence level. The domain structure of all 6 CCN proteins is the same, and with individual domains, there is substantial conservation of amino acids, including a motif of 38 cysteine residues which are present within the primary structure of all CCN proteins except CCN6, which lacks four of them. This sequence similarity contributes to a highly conserved primary structure common to all CCN proteins(119). The flexible hinge region between the VWC and TSP domains of CCN proteins shows considerable variation in both composition and length, with CCN1 having a particularly long hinge region (~100 amino acids long), while that of WISP-1/CCN4 is shorter (~26 amino acids in length)(119). While there is some sequence diversity within individual domains, there are several conserved motifs which likely correspond to protein binding sites, including an α5β3 integrin binding site on the VWC domain,

two α6β1 integrin binding sites on the TSP1 domain and the CT domain, and a heparan sulphate proteoglycan binding site on the CT domain. There is a hierarchy of sequence similarity between the different CCN proteins, shown in Figure 1.7(118).

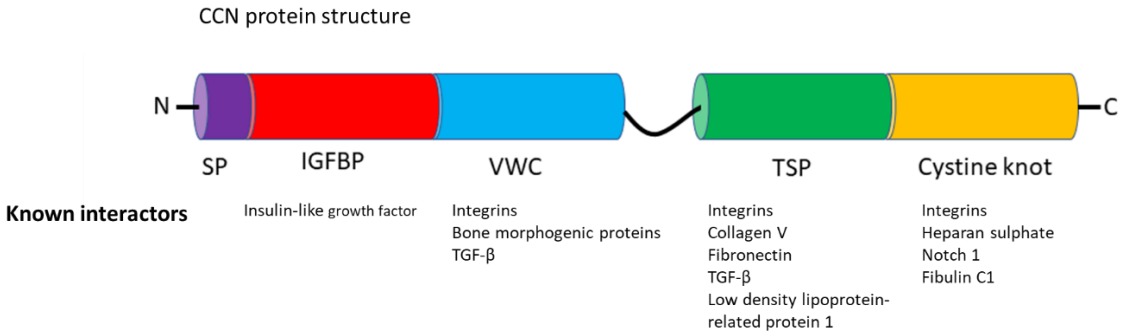


Figure 1.6: Domain structure of CCN proteins. SP is signal peptide, IGFBP is insulin-like growth factor binding domain, VWC is von Willebrand factor type C domain, TSP is thrombospondin response domain. Below each domain is a list of known interaction partners for that domain. Interaction partners taken from(130).

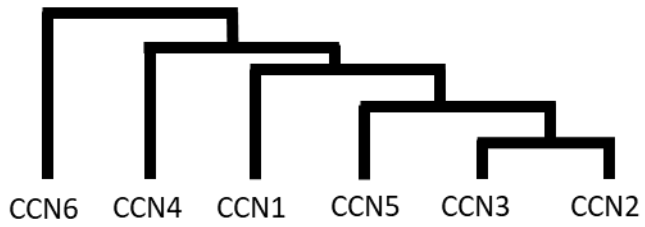


Figure 1.7: Evolutionary hierarchy of CCN proteins, showing their relative sequence similarity via proximity on a dendrogram. Adapted from Nakamura and Bornstein, 2020(118)

Regulation of WISP-1 transcription is reported to be via β-catenin signalling and a Wnt-responsive cyclic AMP response element binding protein (CREBP) element(131). Wnt1 ligand was initially shown to induce WISP-1 expression in mammary epithelial cells(132). A luciferase reporter assay using the WISP-1 promoter showed that Wnt-1 strongly induced WISP1 expression, and that this was mediated by β-catenin signalling. However, this is not mediated via TCF/LEF, the most common transcription factors associated with the activation of Wnt-responsive genes. Instead, a cyclic AMP response element binding protein (CREB) responsive element in the promoter sequence is critical for WISP-1 expression(104).

WISP-1 mediates its functionality via protein-protein interactions, in manner similar to other CCN proteins. As well as interactions with integrins detailed below, a large-scale co-

immunoprecipitation (co-IP) study of multiple potential therapeutic targets in cancer, including WISP-1, identified WISP-1 as interacting with the EMT-associated protein zinc-finger E-box binding protein 2 (ZEB2) and E-cadherin (CDH1)(133).

1.21 WISP-1 in acute lung injury

Mechanical stretch has also been shown to induce WISP-1 expression in mouse lung tissue(134). This study showed that mechanical stretch of mouse lung tissue can induce EMT of alveolar epithelial type 2 cells. Mechanical stretch occurs during breathing, and can lead to epithelial damage, a process implicated in the onset and pathogenesis of IPF. This study suggested WISP-1 induction was required for AT2 EMT, a process that has been implicated in IPF pathogenesis. This study also suggested that fragments of hyaluronan, an abundant extracellular matrix component, may induce this WISP-1 production in an innate immune-responsive manner, via Toll-like receptor 4(134). Hyaluronan fragments are an important inflammatory marker in bleomycin induced lung injury in mice, the principal model system for IPF(135). An association with WISP-1 and increased severity of ventilator induced lung injury was reported using haplotype association mapping, with single nucleotide polymorphisms in the 5' untranslated region of the WISP-1 gene being enriched in mice with worse lung injury. This study showed WISP-1 production increased in mouse acute lung injury, and co-immunoprecipitated WISP-1 with TLR4, suggesting a potential mechanism of action and a protein interaction partner of WISP-1 in the lungs(136). Another study on WISP-1 in acute lung injury in mice suggested a functional interaction between WISP-1 and integrin β_6. Synthetic RGD peptides, mimicking the integrin receptor site, reduced this interaction(137). Furthermore, Chen et al. showed integrin $\alpha_V \beta_3$ expression increased in mice after acute lung injury, and suggested that WISP-1 $\alpha_V \beta_3$ integrin interaction led to an activation of ERK signalling in macrophages that primed them to increased lipopolysaccharide sensitivity(138).

1.22 WISP-1 in IPF

WISP-1 has shown to be upregulated in IPF – a microarray screen comparing gene expression in IPF lung to normal lung showed an increase in WISP1 levels in comparison to other forms of pulmonary fibrosis(72). Comparison of gene expression signatures in IPF tissue to hypersensitivity pneumonitis showed a strong upregulation of genes associated with known profibrotic processes, including TGF-β signalling, collagen deposition, and MMP activity. As well as WISP1 other significantly upregulated genes included multiple MMP genes, including MMP7, collagens, the Wnt signalling antagonist secreted frizzled related protein 2 (SFRP2) and factors associated with

bone and cartilage development, including asporin and osteoblast specific factor 2. This was contrasted with the gene signatures associated with hypersensitivity pneumonitis, which showed increased expression of proinflammatory factors such as chemokines and interferons. This paper defined a set of genes which are associated with fibrosis, including WISP-1(72).

Mouse models of lung fibrosis have been used to study potential effects of WISP-1 in IPF, with ATII cells from bleomycin treated mouse lung showing an increased expression of WISP-1. Studies of airway remodelling in rats suggest a role for WISP-1 in remodelling, and also suggested that WISP-1 could act to increase fibroblast proliferation and collagen production in IPF(139).

The proliferative activity of WISP1 may be mediated by interleukin 6 (IL-6). A study using primary human lung fibroblasts showed that the presence of WISP-1 is required for TGF-β and TNF-α induced IL-6 production and that this IL-6 is capable of increasing fibroblast proliferation. WISP-1 was also shown to be important for fibrosis, fibroblast proliferation and ECM deposition in heart tissue in myocardial infarction, suggesting that its fibrogenic properties are not unique to IPF(140).

Several studies have tried to clarify the role of WISP-1 in IPF. Königshoff et al. identified elevated WISP-1 expression using a microarray screen in ATII cells isolated from bleomycin-treated mice, compared to ATII cells from control mice. This study also identified increased proliferation of these cells in bleomycin-induced fibrosis, and used immunohistochemical staining of WISP-1 to identify increased WISP-1 staining in bleomycin mouse lung tissue Several Wnt-responsive genes were also upregulated in bleomycin-treated mouse lungs, including the CCN protein CTGF. WISP1 expression was also identified in cultured human ATII cells, and WISP-1 protein was localised to ATII cells in human IPF tissue. Functional investigation of cultured ATII cells identified increased proliferation in cultured mouse ATII cells stimulated with WISP-1, and an increase in the expression of profibrotic and EMT-associated genes. Finally, this study showed that administration of an anti-WISP-1 neutralising antibody attenuated fibrosis in bleomycin-treated mice(141).

A second study by the same group identified WISP-1 as being expressed in lung fibroblasts, rather than ATII cells, however. TGF-β treatment was shown to induce WISP-1 expression in primary human lung fibroblasts, and this induction was identified as being partially regulated by a microRNA, mir92A(142). However, other studies have called this induction of WISP-1 expression by TGF-β into question, suggesting that increased TGF-β levels lead to a decline in WISP-1 production by lung fibroblasts(143).

In summary, WISP-1 production is increased in IPF, and it has been shown to mediate various profibrotic effects, including ATII cell proliferation, EMT, and profibrotic gene expression. However, the location of its expression in the IPF lung is ambiguous, with separate studies from the same group identifying WISP-1 expression in both ATII cells and fibroblasts in fibrotic lung tissue.

1.23 WISP-1 in other fibrotic diseases

WISP1 has also been identified as playing a role in the pathogenesis of other fibrotic diseases. WISP-1 expression is increased in an experimental liver fibrosis model(144). WISP-1 expression was also associated with an increase in TGF-β-induced fibrosis in renal fibrosis(145). This suggests that any profibrotic role mediated by WISP-1 in IPF may have analogues in fibrotic diseases in other tissue types.

1.24 WISP-1 in cancer

WISP-1 has also been shown to have a role in cancer. Early studies showed an increase in tumourigenesis in cells expressing WISP-1 injected into mice, showing an oncogenic effect of WISP-1 in certain cancers(131). Expression levels of WISP1 RNA are also increased in multiple types of cancer, including breast, lung, colorectal and gastric cancers(146). WISP-1 acts as an oncogene in human breast cancer. WISP1 mRNA is upregulated in breast cancer, and treatment of MCF-7 cells, a breast cancer cell line, with recombinant WISP-1 led to increased cellular proliferation, and tellingly, induction of epithelial to mesenchymal transition (EMT)(147).

A study in pancreatic carcinogenesis have shown that mutation in P53, a major tumour suppressor protein leads to decreased ubiquitination and degradation of WISP-1. Elevated WISP-1 expression led to worsened prognosis in pancreatic carcinoma patients regardless of P53 mutation status. The same study also explored the role of WISP-1 in cellular malignancy, showing that elevated levels of WISP-1 production using a lentiviral overexpression vector in pancreatic cell lines led to an increase in cellular invasion and tumourigenicity(148).

1.25 WISP-1 in bone development

One of the best-characterised functions for WISP-1 is in driving bone development. WISP-1 expression is increased during bone development and repair, and its expression in differentiating osteoblasts leads to an increase in bone morphogenic protein 2 (BMP-2) induced differentiation. It also had effects on chondrocytes transfected with constitutively-expressing WISP-1, increasing their proliferation, but reducing their ability to differentiate(149). In a different study, bone

development was partially inhibited in WISP1-null mice, with reduced bone mineralisation observed. This was attributed to defective differentiation of bone producing cells(150). WISP-1 signalling has also been implicated in arthritis, with WISP1 gene expression identified as being elevated in osteoarthritis(151).

1.26 WISP-1 in adipose tissue

WISP-1 has also been identified as an inhibitor of adipocyte differentiation. It has been shown to suppress peroxisome proliferator-activated receptor (PPARγ) activity in adipocytes(152). PPARγ drives adipocyte differentiation in adipocytes, and has also been implicated in pulmonary lipofibroblast function in rodents(153). The ability of WISP-1 to inhibit adipocyte differentiation has also been characterised in perivascular stem cells, where WISP1 overexpression was shown to drive these cells towards a bone-type cellular phenotype. WISP1 knockdown in these cells led to an increase in PPARγ expression, and a decrease in the expression of RUNX2, a gene associated with bone development(154,155).

1.27 WISP-1 splice variants

WISP-1 has multiple splice variants, which are implicated in the pathogenesis of several diseases, most notably different cancers, including scirrhous gastric carcinoma and cholangiocarcinoma (156–158). It is unknown what effect these splice variants have on the pathogenesis of IPF, but further investigation of which splice variants are overrepresented in IPF will give rise to a better understanding of how the disease works.

The National Centre for Biotechnology Information (NCBI) describes four principal splice variants of WISP-1 which lead to protein being produced, plus one isoform which undergoes nonsense-mediated decay, but is expressed(156). The four translated variants lack various domains, which may affect the proteins' ability to interact with other proteins. The exons of the WISP-1 gene are largely domain-specific, with one exon corresponding to one protein domain except for a flexible linker between the von Willebrand factor type C domain and the thrombospondin type I repeat(121). Figure 1.8 shows the structure of the WISP1 genomic locus, showing a very large intronic region between the first exon and the second exon. Figure 1.9 shows the domain structure of the four different WISP-1 splice variants, including full-length WISP-1.

Figure 1.8: The structure of the WISP1 genomic locus, showing the 5 exons present in the full-length WISP1 transcript. Source: NCBI/Ensembl.

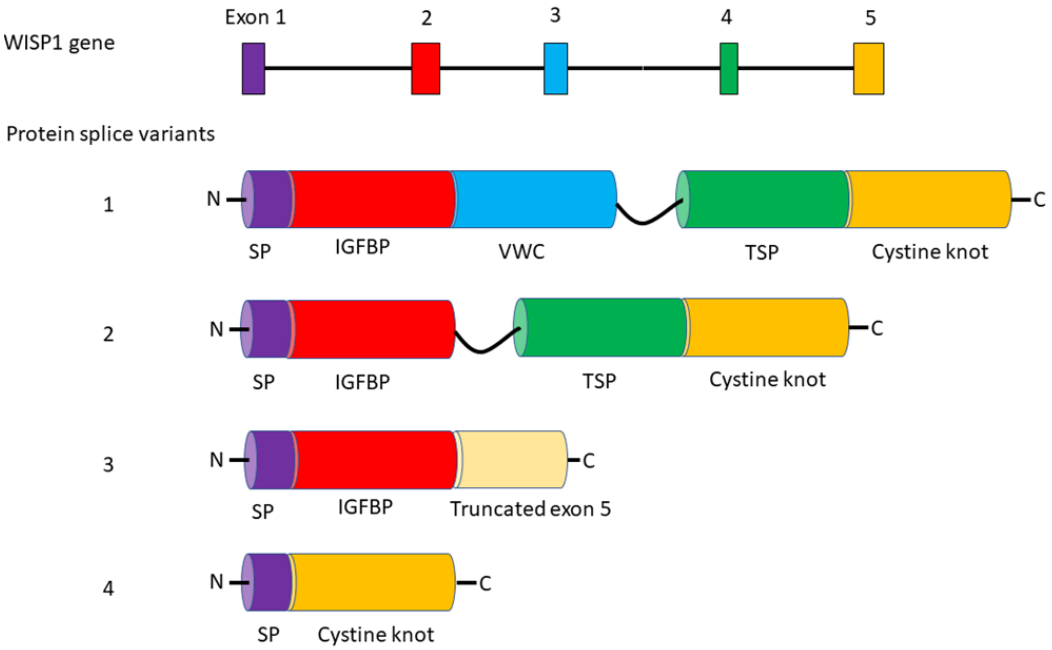


Figure 1.9: WISP-1 splice variants, showing the different domains present in each variant. SP is signal peptide, IGFBP is insulin-like growth factor binding domain, VWC is Von Willebrand factor type C domain, TSP is thrombospondin response domain.

Splice variants of WISP-1 have been described in various diseases. The first alternatively spliced variant of WISP-1 was identified in gastric carcinoma cells, where WISP1v was identified as lacking a von Willebrand factor type C domain and was shown to be expressed in the fibroblast-rich stroma of gastric carcinomas(157). The variant was associated with an increase of cell migration and invasion when transfected into cultured gastric carcinoma cells, an effect not seen when this experiment was repeated with the full-length WISP-1 gene. This variant is identified as WISP-1 variant 2 in this dissertation, consistent with the NCBI naming scheme(156).

Another alternative splice variant of WISP-1 was identified in a hepatocellular cancer cell line(159). They identified a variant containing only the signal peptide and IGFBP domain with a truncated c-terminus caused by a frame-shift splice variant, identified as transcript variant 3 by the NCBI. This paper also identified the previously described variant 2 as being present in hepatocellular cancer cells. Another splice variant, variant 4, lacking all domains but the signal peptide and cystine knot domain, has also been detected in large scale mRNA screens of

pancreatic cancer, but it is not known if this domain has any pathogenic significance(160). Another splice variant detailed on the NCBI website, designated variant 5, is predicted to undergo nonsense-mediated decay, has not been reported to have any pathogenic significance and its expression has not been reported in the literature.

The identification of multiple splice variants of WISP-1, and especially the association of variants 2 and 3 with proliferation and invasion of cancer cells, and the expression of variant 2 in the fibroblast-rich stroma of scirrhous gastric carcinoma(157), suggests that investigation of WISP-1 splice variants may be important in the pathogenesis of IPF, a disease that is characterised by fibroblast proliferation and invasion leading to the formation of fibroblastic foci.

The discrete, regular domain structure of CCN proteins including WISP-1 combined with the large number of protein-protein interactions ascribed to this protein group suggests that interactions may be either domain-specific or require the presence of several specific domains. Splice variants lacking particular domains may therefore lack the ability to undergo regulatory or functional interactions, potentially leading to a role in pathogenesis. A previous PhD study looking at the expression of different splice variants in fibroblast cell lines found that all four splice variants were present both in the cultured alveolar cell line A549 and primary parenchymal fibroblasts(143). Treatment with the profibrotic cytokines TGF-β and TNF-α showed a bell-shaped dose-response curve to both cytokines, with WISP-1 expression increasing initially, and then decreasing with increasing dosage of TGF-β or TNF-α. This study also showed that after 72 hours in culture, primary parenchymal fibroblasts from IPF patients expressed all four splice variants at a greater level than in fibroblasts from healthy controls, suggesting that an upregulation of WISP-1 expression may lead to upregulation of all splice variants. Splice variants 2 and 4 were upregulated in IPF fibroblasts stimulated by TGF-β1, an effect not seen in control fibroblasts, which only demonstrated a minor increase in splice variant 4(143).

1.28 Summary

In summary, IPF is a debilitating and eventually fatal interstitial lung disease associated with aging and characterised by progressive lung fibrosis. Features of IPF include destruction of lung architecture and deposition of stiffened extracellular matrix into the lung interstitium, reducing lung function. Histologically, IPF is characterised by the presence of usual interstitial pneumonia, where lung tissue consists of areas of patchy fibrosis, containing fibroblastic foci, aggregates of ECM-producing fibroblasts and myofibroblasts which produce fibrotic extracellular matrix. Although the cause of IPF is unknown, there are several environmental risk factors associated with disease development, including cigarette smoking, stone dust and exposure to animal

agriculture. There are also some genetic risk factors for IPF, including variants of the MUC5B gene coding for a mucin protein, and genes associated with telomere elongation. Treatment options are limited to two antifibrotic drugs, pirfenidone and the receptor tyrosine kinase inhibitor nintedanib. Although effective at slowing disease progression, they do not constitute a cure. IPF is thought to arise from repeated microinjuries to lung tissue leading to chronic, pathological activation of wound-healing pathways, which drive fibrosis. ECM in IPF is characterised by increased stiffness driven by the prevalence of bone-type pyridinoline collagen crosslinks.

A key cell type in fibroblastic foci is the myofibroblast, a differentiated cell type characterised by the presence of contractile α-SMA filaments and production of ECM components. Myofibroblast differentiation is driven by the profibrotic cytokine TGF-β. Other signalling pathways upregulated in IPF include HIF and Wnt signalling, as well as mechanosensing caused by increased tissue stiffness and myofibroblast contraction.

Alveolar type II cells are also dysregulated in IPF, undergoing epithelial-mesenchymal transition – expressing fewer epithelial markers in favour of mesenchymal markers.

The CCN proteins are a family of six matricellular proteins which have diverse roles in cell signalling, proliferation, survival and adhesion. CCN proteins mediate their functions through diverse protein-protein interactions. CCN proteins that have previously been implicated in fibrosis include connective tissue growth factor (CTGF) and Wnt-inducible signalling protein 1 (WISP-1). WISP-1 expression is increased in fibrotic lung and has been identified as driving profibrotic phenotypes in ATII cells as well as lung fibroblasts. WISP-1 is responsible for driving osteoblast differentiation and chondrocyte proliferation in bone and cartilage development and has oncogenic activity in cancer. This project aims to explore and clarify the role of WISP-1 in IPF, identifying where it is expressed in IPF tissue, characterising the phenotypes of WISP-1 expressing cells, identifying drivers of WISP-1 expression in these cells, and exploring WISP-1 functionality by identifying its protein interaction partners in appropriate cell lines.

Hypothesis and aims and objectives

1.29 Hypothesis

It is hypothesised that WISP-1 is dysregulated in fibroblastic foci, and that this contributes to IPF pathogenesis by affecting protein-protein interactions.

1.30 Aim 1: Identify where WISP-1 is expressed in IPF

1.30.1 OBJECTIVES

- To identify a set of IPF-specific gene expression signatures from publicly available transcriptomics datasets.
- To use RNAseq data from laser-capture microdissection microscopy of normal and IPF alveolar septae and fibroblastic foci to identify where WISP1 is expressed.
- To identify specific cell types which express WISP1 using IPF single-cell RNAseq expression data.
- To map genetic signatures from those cell types onto LCMD RNAseq data using CibersortX *in silico* cell sorting.
- To characterise the expression context of WISP-1 expressing cell types to gain insight in their functional role.
- To confirm WISP1 localisation using an RNAscope *in-situ* hybridisation approach.

1.31 Aim 2: Identify drivers of WISP-1 in IPF

1.31.1 OBJECTIVES

- Identify gene expression signatures associated with signalling pathways present in WISP1 expressing cells.
- Confirm that known drivers of WISP1 colocalise with WISP1 mRNA using RNAscope *in-situ* hybridisation.
- Drive WISP1-associated signalling pathways in appropriate cell culture models, identifying how they affect WISP1 gene expression using real-time quantitative PCR.
- Confirm these effects on protein using western blot analysis.

1.32 Aim 3: Identify functional roles of WISP-1 in IPF

1.32.1 OBJECTIVES

- Identify functional consequences of WISP-1 knockdown in appropriate cell culture models by quantifying marker gene expression
- Optimise a system to overexpress GFP-tagged WISP-1 in human cells
- Identify subcellular localisation of WISP-1
- Use an affinity-purification mass spectrometry workflow to identify interaction partners with GFP-tagged WISP-1 in appropriate cell lines.
- Confirm protein-protein interactions using western blot analysis and ELISA-based assays.

2 Materials and methods

The sources of all materials are identified in the relevant section of each method.

Bioinformatic analysis

2.1 Differential expression analysis of bulk RNAseq data

Differential expression analysis of bulk RNAseq data was performed using the EdgeR R package(161,162). Raw counts data were ingested into R, and converted into a DGElist object with an added grouping variable corresponding to the sample grouping. Published datasets used for differential expression analysis had already been filtered for low-expressing genes prior to publication, so a filtering step was not included. Data were downloaded from the gene expression omnibus (GEO). Details of datasets are described in chapter 3.

Counts data in the DGElist object were normalised using the calcNormFactors() function in edgeR – this calculates a set of normalisation factors which minimises logFC values between the majority of samples for most genes – reducing the tendency of very high-expressing genes to reduce the counts values of lower-expressing genes due to taking up a large proportion of reads in a sequencing run. The default trimmed mean of M-values (TMM) method was used to perform this normalisation. This method is appropriate when most genes in an experiment are not differentially expressed, as was the case here. The estimateDisp() function was used to estimate dispersion, a measure of variance between replicates. A quasi-likelihood F-test was then run to identify differentially expressed genes. Benjamini-Hochberg multiple test correction was applied, and genes were filtered by the BH q-value – genes with q-values of <0.05 were considered differentially expressed.

2.2 Laser capture microdissection RNAseq data generation

Laser capture microdissection of formalin fixed, paraffin embedded (FFPE) lung tissue sections and RNA isolation was performed by Dr Mark Jones. Sequencing data was processed by Dr Milica Vukmirovic (Yale University).

2.2.1 LASER CAPTURE MICRODISSECTION OF FFPE LUNG TISSUE

Laser capture microdissection was done on formalin fixed, paraffin embedded (FFPE) tissue, with control non-fibrotic lung tissue (n = 10) sampled from macroscopically normal tissue away from tumour sites in lung cancer patients, and IPF lung tissue from patients showing UIP with subsequent IPF diagnosis (n = 10). Fibroblastic foci as well as nearby alveolar septae were

subjected to microdissection, forming two distinct sample groups from the IPF tissue. Patients were age and sex matched. 10µm sections were cut from FFPE tissue blocks and mounted on glass slides. Sections were deparaffinised and rehydrated with xylene and gradated alcohols, stained with paradise PLUS staining reagent (Thermo Fisher, cat no. KIT0311) to identify fibroblastic foci, then dehydrated with gradated alcohols to xylene. LCMD was performed using a Leica Laser Microdissection ASLMD microscope, with regions of interest being selected using Leica microdissection software before being excised into RecoverAll Total Nucleic Acid digestion buffer (Thermo Fisher, cat no. AM1975).

2.2.2 RNA ISOLATION

RNA was isolated using the RNAqueous-Micro Total RNA Isolation kit (Invitrogen, cat no. AM1931). RNA concentration and RNA integrity (RIN) numbers were calculated using a Qubit RNA HS Assay Kit and Qubit Fluorometer. RIN numbers were approximately 2 for most samples. Complementary DNA (cDNA) libraries were generated from 15ng total RNA using the Ion Ampli-Seq-transcriptome human gene expression kit (Ion Torrent, cat no. A26325), before sequencing.

2.2.3 RNA SEQUENCING

Sequencing was performed using an Ion Torrent Proton sequencer. Approximately 20 million single-end reads of roughly 100bp in length were produced per sample. Post processing was performed using Cufflinks, and reads were mapped to the University of California Santa Cruz hg19 human genome before calculation of fragments per kilobase of exon per million mapped reads (FPKM) values.

2.3 Differential expression analysis of microarray and LCMD RNAseq data

The R package Limma was used to perform differential expression analysis of microarray data(163). The lmfit() function was used to fit a linear model to each gene, with a specified design matrix, estimating standard errors and fold changes for each gene. Empirical Bayes smoothing was applied to these standard errors using the eBayes() function. Benjamini-Hochberg multiple test correction was applied as with the EdgeR analysis above, with significantly differentially expressed genes being defined as having a q-value of <0.05. Data were downloaded from the gene expression omnibus (GEO). Details of datasets are described in chapter 3.

This methodology was also used for differential expression analysis of laser-capture microdissection RNAseq data. As raw counts for this data were not available, the normalised, FPKM values for each gene was used for differential expression analysis. These data were

normally distributed, and so the usual, negative binomial methodology for RNAseq data analysis was inappropriate. Limma differential expression analysis was used instead.

2.4 Data processing and visualisation

Data processing and visualisation was done using the R programming language(164). The R package ggplot2 was used to generate PCA plots and heatmaps(165). Gene ontologies and pathways were identified by inputting gene lists to GOrilla (Gene Ontology enRIchment anaLysis and visuaLizAtion tool http://cbl-gorilla.cs.technion.ac.il/)(166,167). A list of target genes from differential expression analysis was input into GOrilla, along with a list of background genes (i.e., all genes quantified in a dataset). When a background list of genes was unavailable (e.g., for interaction proteomic data), a ranked list of gene names was used for the input instead. R code for gene ontology treemaps was generated using GOrilla, downloaded, and modified to produce better quality plots.

2.5 Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed using the Broad Institute's dedicated GSEA software (downloadable at http://www.gsea-msigdb.org/gsea/index.jsp)(168). This takes two user-defined classes of genes and ranks them according to the correlation between their expression level and the class definition. Genes with high expression in the first class and low expression in the second will have a higher rank, and vice-versa. Genes with little difference between classes will be mid-ranked. Gene set enrichment analysis is then performed on this ranked list of genes to determine which pre-defined gene sets are enriched at the top and bottom of this ranked list. Firstly, an enrichment score (ES) is calculated. Genes are weighted by their position in the ranked list. The ES is calculated by walking down the ranked gene list - it is increased when a gene present in a gene set is reached, and decreased when a gene not present in the set is found. By plotting the enrichment score going down the gene list, the distribution of a gene set is calculated. This distribution is dependent on the pattern of expression across the whole dataset, with positively enriched gene set having their distribution skewed heavily towards the top ranked genes, and vice-versa for negatively enriched gene sets. If a gene set is distributed evenly across the ranked list of genes, it will have a flat distribution profile, not heavily weighted to either side. In order to determine the significance of these gene sets, the class labels are rearranged and a null distribution for the gene set is calculated; the ranked gene list and the enrichment scores are recalculated for this null distribution, and significance is determined by comparing the original ES distribution with the null ES distribution. If multiple gene sets are used,

the p-values are adjusted to account for multiple hypothesis testing. Enrichment scores are normalised based on the gene set size, and a false discovery rate (FDR) is calculated on the normalised enrichment scores.

A variety of gene sets were used for the analysis, with different databases of gene sets being used for different aims. Gene ontologies, Kyoto encyclopedia of genes and genomes (KEGG) pathways, and Hallmark gene sets were used(168–171). Gene ontologies are curated gene sets that contain the genes associated with molecular functions, cellular components, and biological processes. The genes associated with a given GO term are therefore functionally or spatially related, but not necessarily coregulated or involved in the same biological pathways. GO terms are also nested, with broader terms encompassing multiple more specific terms. A gene can be referenced by multiple GO terms. KEGG pathways, by contrast, specifically encompasses biological pathways, i.e. gene sets which are involved in the same molecular pathway. Path- ways as defined by the KEGG are molecular interaction or reaction networks. Unlike the sets defined by GO terms, the members of a KEGG pathway set all contribute to the same process, for example they are all members of the same signalling cascade, or they all contribute to the same metabolic pathway(169). Finally, hallmark gene sets are highly curated gene sets taken from the molecular signatures database (MsigDB)(168). They represent gene coexpression networks which are associated with particular biological states or processes. The hallmark gene sets are refined from multiple other gene sets, preserving the information from those sets while reducing redundancy and variation. The molecular signatures database was originally developed for use with GSEA(168).

2.6 Single Cell RNAseq data analysis

Single cell RNAseq counts data, along with associated metadata, was downloaded from the Gene Expression Omnibus (GEO) database (accession number GSE135893)(58). Counts data was processed using the Seurat R package(172). Data processing unless otherwise stated was done in R 3.6 running on Windows 10. The computer used for this analysis had an Intel Core I7 7700 CPU and 16GB of RAM.

2.6.1 SINGLE CELL DATA FORMATTING AND INGEST

Counts data was downloaded as a .mtx (sparse matrix) file from GEO. The nature of single cell RNAseq data – large dimensionality, but most counts values for a given cell being 0 – make sparse matrices ideal for dealing with otherwise prohibitively large data. The .mtx file was supplemented

by separate .txt file containing gene names and unique cell identifiers. Associated metadata was included in a separate .csv file.

Single cell counts data was uploaded into R using the readMM() function from the Matrix R package(173). Unique cell identifiers and gene names were uploaded separately and assigned as column and row names to this matrix. Metadata was uploaded as a separate data frame, and cells which had failed the quality control requirements detailed in Habermann et al. were filtered out, leaving only the cells from the original paper.

2.6.2 SEURAT ANALYSIS

Once processed and formatted in R, counts data and metadata were incorporated into a Seurat object, normalised using Seurat's default LogNormalise method. This normalizes each gene expression measurements for each cell by the total gene expression, multiplies this by a scale factor (10,000 by default), and log-transforms the result.

Highly variable genes were then identified using the FindVariableFeatures() function. This streamlines downstream analysis by identifying the most variable features in the dataset. As these are the source of most of the variation in the dataset, subsequent dimensional reduction on just these features significantly reduces compute requirements while still capturing the variation in the dataset.

Principal components analysis was then performed on these features. Dimensional loadings were visualised and heatmaps plotted for the top variable genes associated with each principal component. In order to determine the dimensionality of the dataset (i.e. how many principal components to use for subsequent analyses), a JackStraw plot was generated with 100 replicates, and an appropriate number of principal components was selected based on the change in the distribution of P-values for each principal component. This was confirmed by looking for an "elbow" in the proportion of variation explained in an elbow plot of the first 20 principal components. For this dataset, there was a clear elbow at approximately 15 principal components, so 15 PCs were used for subsequent analyses. However, t-SNE and UMAP analysis was performed using 10 – 30 PCs, and there was little visual difference in the plots generated.

Clusters were assigned based on their published IDs – they were not determined *de novo.* However, a test clustering was performed on fibroblasts alone and largely matched the clustering performed by the original authors. This was taken as sufficient validation that the originally published cluster assignments could be used.

Cells identified as fibroblasts (i.e. mesenchymal cells which were not smooth muscle cells) in the published dataset were isolated and the above Seurat analysis was performed on these cells alone as well.

2.6.3 IDENTIFICATION OF MARKER GENES

Marker genes for individual clusters were identified using the Seurat FindMarkers() function. This identifies genes which are differentially expressed in a cluster, either relative to the rest the dataset, or between specified individual clusters. The default Wilcoxon rank sum test was used for this, with a minimum percentage cell expressing a given gene of 10%, and with a minimum log fold change threshold of 0.25.

2.6.4 T-STOCHASTIC NEAREST NEIGHBOUR EMBEDDING (T-SNE) AND UNIFORM MANIFOLD APPROXIMATION AND PROJECTION (UMAP)

Although principal component analysis is an effective way to capture variation in a dataset, it does not sufficiently capture the non-linear nature of single cell RNAseq data, and thus PCA plots do not recapitulate clusters identified by other means, including the hierarchical clustering analysis from the original paper. Therefore, to preserve the structure of the data in 2 dimensions, it is necessary to perform other dimensional reduction techniques for adequate data visualisation. Two-dimensional reduction techniques were used – t-stochastic nearest neighbour embedding (t-SNE) and uniform manifold approximation and projection (UMAP). Both of these preserve the complexity of the dataset in fewer dimensions, and are suitable for very high dimensional data such as single cell data(174,175).

For the whole dataset, this proved too computationally expensive to run on a standard computer, with excessively high RAM requirements, so t-SNE and UMAP analysis was performed in an R instance running on the University of Southampton's IRIDIS 4 high performance compute cluster, on a compute node with 64GB of memory. 15 principal components were used as the input for this analysis. Once run, the t-SNE and UMAP embeddings (i.e. (x,y) coordinates) were saved as a csv file, uploaded into R, and high quality plots using these coordinates and associated metadata were generated using ggplot2.

2.7 CibersortX analysis

CibersortX analysis was performed using the CibersortX web portal (https://cibersortx.stanford.edu/), and was used to derive gene expression signatures associated with different cell types identified in the Kropski single cell RNAseq dataset. These signatures

were then mapped onto LCMD RNAseq data to quantify the proportion of each cell type present in individual LCMD tissue samples – a “virtual cell sorting” methodology(176,177).

2.7.1 SIGNATURE MATRIX CREATION

Single cell RNAseq counts data from Habermann et al. was loaded into R using the sparse matrix methodology detailed above. 90 cells from each cell type were then randomly selected from cells from control and IPF tissue using the sample() function in R, with a seed of 2. These were combined into a smaller (no longer sparse) matrix, with column names corresponding to cell type, and row names corresponding to genes. Genes with no expression across the dataset were filtered out, and this matrix was saved as a tab-delimited .txt file. This was uploaded to the CibersortX web portal. Reduction of the dataset in this way was required due to inherent processing limitations for large datasets within CibersortX – 90 cells per cell type was determined to be the maximum for signature matrix generation for this dataset. These data were then used to produce a signature matrix, identifying expression signatures of up to 300 genes per cell type. Signature matrix generation setting were left at default, apart from the min. expression value, which was set to 0 to account for the low counts values associated with the 10X Chromium single cell sequencing methodology used for this dataset.

2.7.2 CELL FRACTION IMPUTATION

The signature matrix above was then used impute cell fractions using the Cibersort algorithm. The LCMD RNAseq data was used as the mixture file for cell fraction determination. Quantile normalisation was disabled for this, and 100 permutations were performed. Once cell fractions had been imputed, this data was downloaded and graphed in R using ggplot2.

2.8 Gene set variance analysis

In order to identify enrichment of gene sets associated with certain processes in the single-cell RNAseq dataset, gene set variance analysis was performed. This was done using the GSVA R package. Gene sets were curated from a variety of sources, from gene ontology lists to a published 15-gene signature associated with HIF signalling(178). Gene sets were uploaded to R and converted to entries in a list object. Single cell data for all fibroblast subtypes were also uploaded, and The gsva() function from the GSVA package was used to calculated enrichment scores for each cell in this dataset using the different gene sets as the inputs. This uses the GSVA method of Hänzelmann et al. to calculate an enrichment score for each sample (in this case, individual fibroblasts)(179). These scores were then plotted onto previously calculated t-SNE coordinates using a diverging colour scale centred on 0 using a ggplot2 scatterplot.

2.9 Pseudotime analysis

Pseudotime analysis was done using the monocle3 R package(180). Single-cell RNAseq data was loaded into R, and converted to a monocle cell dataset using the new_cell_data_set() function. Pre-processing consisted of Principal components analysis and dimensional reduction using UMAP, using the same methodology as Seurat. Once UMAP embeddings were calculated, pseudotime analysis was performed. First, the learn_graph() function was used to calculate a principal graph – this displays the paths of variation within the UMAP embeddings – i.e. the branched structure of the underlying data. These trajectories were then used to calculate pseudotime scores for individual cells, based on where they fit on this graph. Pseudotime scores were plotted based on selection of individual nodes on the graph – nodes were selected based on their placement at one end of the graph, and then pseudotime scores were plotted as a colour scale based on their distance from that node. Pseudotime analysis provides an insight into how cells are related to each other.

2.10 Code availability

All R code used for Bioinformatic analyses is available on GitHub in the github repository https://github.com/JoeBell1/repo.git .

Molecular Biology

2.11 Generation of eGFP-tagged WISP-1 constructs

eGFP-tagged WISP-1 expression constructs were generated for affinity purification mass spectrometry. Polymerase chain reaction (PCR) amplification and template sources PCR was performed using hot start high-fidelity DNA polymerase KOD, from Thermophilus kodakarensis. PCR primers were ordered from Invitrogen. WISP-1 variant 1 template DNA was sourced from A549 cDNAs in which the expression of full-length WISP-1 had been previously verified. Variants 2, 3 and 4 had previously been cloned into the pcDNA3.1(+) vectors by Dr Lyndsy Ambler and Dr David Smart and the resulting construct sequenced, and these were used as the template for PCR amplification of those variants(143). The pcDNA3-EGFP vector (Addgene, cat. No. 13031) was used as the template for enhanced green fluorescent protein (eGFP).

2.11.1 VECTOR

The pcDNA 3.1(+) vector was used as the vector for insertion of eGFP-tagged WISP-1. This vector contains ampicillin and kanamycin resistance genes, allowing selection of bacteria containing the

plasmid using ampicillin containing agar, a cytomegalovirus (CMV) promoter and polyA sequence flanking a multiple cloning site containing 11 restriction endonuclease restriction sites (Figure 2.1). pcDNA 3.1(+) is a well characterised mammalian expression vector which shows strong expression in a variety of cell lines.

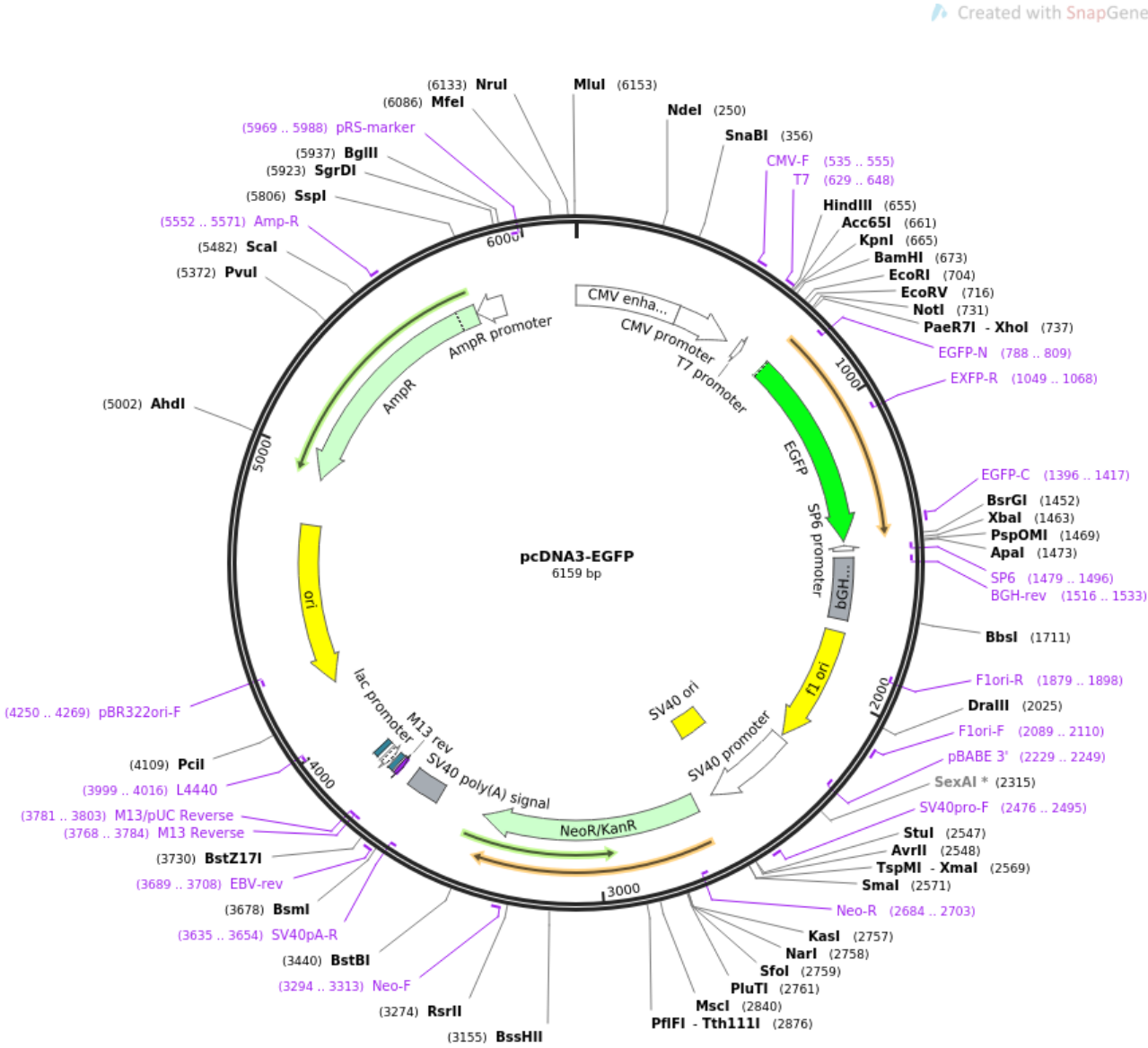


Figure 2.1: Plasmid map of pcDNA3.1(+). Key features, including antibiotic resistance gene, multiclonal site, promoter and poly(A) sequence are shown.

2.11.2 PRIMER DESIGN

PCR primers were designed to amplify the full length of the WISP-1 variants. Primers were designed with unique restriction sites present in the multiple cloning site of pcDNA 3.1(+). The 5' primers for amplification of all WISP-1 variants and 3' primers for amplification of eGFP both contain these restriction sites to allow for digestion and annealing of the WISP-1-GFP assembly. 3'

primers for the WISP-1 PCR and 5' primers for the GFP PCR were designed with compatible ends that overlap in sequence, allowing the annealing and amplification of the two PCR products to produce DNA containing the WISP-1 variants in-frame with the eGFP gene. The resulting overhangs overlap the 5' end of the GFP gene sequence and the 3' end of the WISP-1 sequences, so they could be used for fusion PCR using the overlap as a primer to give a DNA molecule with both the WISP-1 GFP present. KOD hot-start DNA polymerase (Merck, cat no. 71085) was used to generate DNA constructs. PCR reactions were performed in a BioRad T100 thermocycler. 200ng of cDNA or 50ng of plasmid DNA were combined with 1µl of KOD polymerase, 25µl KOD reaction buffer, 15pmol each of forward and reverse primers, and made up to 50µl with ddH_2O. Reactions were heated to 94°C for 2 minutes, before 10 seconds at 98°C to denature DNA strands, followed by 68°C for 1 minute to anneal and extend primers. The latter two steps were repeated 30x. Table 2.1 shows the sequences of all the primers used for these amplification and fusion PCR reactions. Primers were supplied by Invitrogen. Plasmids and PCR products were run on a 2% agarose (Sigma-Aldrich, cat no. A9539) gel in Tris-acetate-EDTA buffer (242g tris-base, 57.1ml glacial acetic acid, 100ml 0.5M EDTA (Sigma-Aldrich, cat nos. T1503, A6283, E9884)). Purification of PCR products was performed using Qiagen's QIAquick gel extraction kit (cat no. 28704).

Table 2.1: Invitrogen primers used for amplification of WISP-1 and GFP genes and creation of fusion PCR constructs

Primer name	Sequence (5' – 3')
WISP-1 forward	TTGCTAGCATGAGGTGGTTCCTGCCCTGG
WISP-1 reverse (no tag)	TTGGTACCGCCTAGTTGGCAATTTCTGAGAAG
WISP-1 v124 overhang reverse	CCTCGCCCTTGCTCACCATGTTGGCAATTTCTGAGAAGTCAG
WISP-1 v3 overhang reverse	CCTCGCCCTTGCTCACCATTTGTCCATGCAAACTCCACAGTACTTG
GFP v124 overhang forward	CTGACTTCTCAGAAATTGCCAACATGGTGAGCAAGGGCGAGG
GFP v3 overhang forward	CAAGTACTGTGGAGTTTGCATGGACAAATGGTGAGCAAGGGCGAGG
GFP reverse	TTTGGTACCTTACTTGTACAGCTCGTCCATGCCGA

2.11.3CLONING

WISP-1 GFP fusion PCR products were ligated using restriction digestion and ligation. 1µl (10u) of restriction endonucleases Kpn1 and Nhe1 (NEB, cat no. R0142S and R0131S). NEB buffer 2.1 was used for digestion in 50µl reaction volume. were used for insertion into the multiple cloning site of pcDNA3.1(+). 1µg Plasmid was digested for 2 hours at 37 °C. T4 DNA ligase (NEB, M0202S) was used for the ligation of DNA fragments. 50ng of gel-purified, restriction-digested plasmid was

combined with 30ng of purified, digested insert in T4 reaction mix (1µl enzyme, 2µl 10x reaction buffer, made up to 20µl with ddH_2O) and incubated at room temperature for 2 hours. DNA was transformed into TOP10 chemically competent E. coli cells (Invitrogen, cat no. C4040-10). 50µl of bacteria were thawed on ice and 4µl of ligation reaction was added to the cells. This was incubated for 30 minutes on ice before a 40 second heat shock at 42 ◦C. 350µl SOC high efficiency transformation media (Thermo Fisher, cat no. 15544034) was added to the cells and the mixture was incubated at 37 ◦C for 15 minutes before being plated onto ampicillin agar overnight for selection. Colonies were picked using a pipette tip, then added to 3ml of ampicillin-containing (50µg/ml) Luria-Bertani medium (LB, Thermo Fisher, cat no. 12780029) for overnight growth. Plasmid DNA was purified using a miniprep kit (Qiagen, cat no. 27104).

Plasmids were assessed by diagnostic digestion. They were incubated with Nhe1 and Kpn1 as above for two hours before being run on an agarose gel and DNA fragments visually inspected. If the correct size fragments were present, plasmids were sent for sequencing (Source Bioscience), and once correct sequences were found, large amounts (approximately 300µg) of transfection-grade plasmid were obtained using Qiagen's maxiprep kit (Qiagen, cat no. 10023). All DNA samples were quantified using the NanoDrop 2000 spectrophotometer (Thermo Fisher, cat no. ND-2000), using 260nm absorbance to quantify DNA amounts. Purity was assessed using the ratio of 260/230nm absorbance measurements, with a higher ratio (approximately 2.0 and above) indicative of high sample purity.

Cell Biology

2.12 Cell culture

2.12.1 CELL CULTURE MEDIUM

Fibroblast cell lines were cultured in complete fibroblast culture medium, consisting of Dulbecco's modified Eagle's medium (DMEM) (Gibco, cat no. 11960 044) containing 10% foetal bovine serum (FBS, Gibco, cat no. A31605). Medium was supplemented with2mM L-glutamine 1mM sodium pyruvate, 50 units/ml penicillin/50µg/ml streptomycin, and 1x non-essential amino acids (Thermo Scientific, cat nos. 25030024, 11360039, 15140122, 11140035). All cell culture was done under aseptic conditions in a microbiological safety cabinet (Herasafe KS, cat no. 51027118)

2.12.2 THAWING OF CELL LINES

HEK293T, HeLa and MRC-5 cell lines were removed from liquid nitrogen storage. 1ml of cell culture medium, prewarmed to 37 ◦C was added to cryovials, quickly thawing the cells. Cells were

added to T75 flasks with 10ml of culture medium and incubated at 37 °C/5% CO_2 overnight. Medium was exchanged after overnight incubation to remove traces of DMSO cryoprotectant from the cultures.

2.12.3 SUBCULTURING OF CELL LINES

Cells were routinely passaged in T75 flasks (from?). Medium was removed and cells were washed with sterile PBS and detached from the flask by trypsinisation with 1ml trypsin solution (trypsin-EDTA 0.5% diluted 1:10 with Ca^{2+}/Mg^{2+} free Hank's balanced salt solution, Thermo Fisher, cat no. 15400054, 14170120). Trypsin was pipetted off and cells were incubated at 37 °C until they had all detached from the flask surface (approximately 5 min for MRC-5 cells, 3 min for HEK293T and HeLa cells). 1ml of media was added to halt trypsinisation. Trypsinised cells were then split at an appropriate ratio – 1:5 to 1:8 for HEK293T and HeLa cells, 1:2 to 1:4 for MRC-5 cells. The appropriate volume of cell suspension was added to fresh T75 flasks, and medium was added to 12ml total volume.

2.12.4 PRIMARY FIBROBLAST CULTURE

Primary IPF and control fibroblast cultures were performed by Dr Lizzie Davies or Dr Chris Brereton. All human lung experiments were approved by the Southampton and South West Hampshire and the Mid and South Buckinghamshire Local Research Ethics Committees (ref 07/H0607/73), and all subjects gave written informed consent.

Primary pulmonary fibroblasts were isolated by taking surgical lung biopsies and cutting them into approximately 2mm x 2mm pieces. Pieces were transferred to a 6-well plate and scratched into the surface of the plate using a disposable scalpel, in order to encourage outgrowth of fibroblasts. 3ml of complete fibroblast culture medium was added to the wells, followed by incubation at 37oC/5% CO2. Empty wells were filled with Hank's Balanced Salt Solution (HBSS, Gibco, cat no. 24020117) to keep the plate humidified. Cells were incubated for 7 - 21 days, with media being replaced every 2 days, until 70-80% confluence. Once this confluence was reached, cells were passaged once before freezing as above. Cells were checked for mycoplasma contamination before use. Isolation of the primary pulmonary fibroblasts used in this study was conducted by Dr Franco Conforti.

2.12.5 CULTURING OF ATII ER: KRASV12 CELL LINE

The ATII ER: KRASV12 cell line was generated from primary cultured alveolar type II cells by Olivier Pardo and Michael Seckl (Imperial College, London)(181,182). ATII ER: KRASV12 are immortalised ATII cells containing KRAS(G12V) fused to the estrogen receptor ligand-binding domain that is

conditionally responsive to 4-hydroxytamoxifen (OHT). In the absence of hormone, the fusion protein is held in an inactive state, probably due to complex formation with heat shock protein 90 (HSP90). Addition of hormone causes a conformational change that dissociates HSP90, resulting in the rapid activation of the fusion protein. These cells were cultured in DCCM-1 media (Geneflow, cat no. K1-0502) containing 10% FBS. Cell culture and mRNA isolation from them was performed as below by Liudi Yao.

2.12.6 CELL COUNTING

Cells were counted using a haemocytometer. 10µl of a 1:1 ratio of cell suspension : trypan blue dead cell stain was added to the counting chamber. Live cells were counted in four $1mm^2$ squares. Cell suspension was diluted using additional media for large numbers of cells, to give more accurate cell counts. Cell counts were calculated using the equation:

$$\frac{Cells}{ml} = \frac{Cell\ count\ \times 2 \times 10^4}{4} = \frac{Cell\ count\ \times 10^4}{2}$$

Where 2 is the dilution factor, and 4 is the number of squares counted.

Cells were plated according to the cell type and the surface area of the plate used. For a 6 well plate, HEK293T cells were plated at a density of 500,000 cells/well, whereas MRC-5 cells were plated at a density of 100,000 cells/well, to take into account the larger surface area covered by individual MRC-5 cells. For differently sized plates or dishes cell numbers were scaled up or down accordingly. Transfection of cells

HEK293T cells were transfected using TransIT-2020 transfection reagent (Mirus, cat no. MIR5404). Cells were plated at a density of 500,000 cells/well in 6 well plates 24 hours before transfection. Cells were transfected with 2.5g DNA. DNA was added to 250µl opti-MEM reduced serum media (ThermoFisher, cat no. 31985062), then 7.5µl of transfection reagent was added. Transfection mixture was incubated for 20 minutes at room temperature then pipetted dropwise onto each well, containing 2ml complete growth medium (2ml total media volume).

HeLa cells were transfected using lipofectamine 2000 transfection reagent (ThermoFisher, cat no. 11668027). They were plated at 100,000 cells/well in 6 well plates and transfected with 2.5µg DNA. 250µl opti-MEM was mixed with DNA, and in a separate tube 250µl opti-MEM was mixed with 7.5µl lipofectamine. This was pipetted onto cells and incubated for 24 hours.

MRC-5 cells were transfected using TransIT-X2 reagent (Mirus, cat no. MIR6003) at a 3:1 Transfection reagent:DNA ratio. 7.5µl transfection reagent and 2.5µg DNA were diluted in 250µl

opti-MEM per well in a 6-well plate containing 300,000 cells and 2ml medium per well. Transfection mixtures were incubated for 30min at room temperature to allow complexes to form, before dropwise addition to cells.

2.12.7 VERIFICATION OF TRANSFECTION

Constructs containing GFP were verified as having been successfully transfected using fluorescence microscopy (Leica CTR7000).

2.13 Confocal microscopy

Confocal microscopy was performed by Dr David Johnston at the Biomedical Imaging Unit, University Hospital Southampton, on a Leica TCS-SP8 Laser Scanning Confocal Microscope. Images were processed using Leica Application Suite X (LAS-X) software. Z-stacks were taken and images selected from the Z-stacks.

2.14 GFP-trap co-immunoprecipitation

The GFP-trap-agarose kit (Chromotek, cat no. gtak-20) was used for immunoprecipitation of GFP and GFP-tagged WISP-1 from cell lysates. Cells were seeded in 6-well plates the day before transfection as above. All 6 wells of one 6-well plate was used per immunoprecipitation reaction.

48 hours post-transfection, cells were harvested using a cell scraper. Cells were collected in a microcentrifuge tube, spun at 300g for 5 minutes to pellet, then washed 3x with PBS, spinning down each time. Cells were then lysed in 200µl ice-cold GFP-trap lysis buffer (in kit, 10 mM Tris/Cl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5 % Nonidet™ P40 Substitute, 0.09 % sodium azide) supplemented with Halt protease inhibitor cocktail (Thermo Fisher, cat no. 87786) for 30 minutes on ice, pipetting extensively every 10 minutes. Lysates were spun at 17,000xg in a microcentrifuge at 4^{o}C for 10 minutes to remove any cellular debris and the supernatants transferred to a fresh microcentrifuge tube. Lysates were then either frozen at -80^{o}C for later use or used immediately for Co-immunoprecipitation.

GFP-Trap A beads were resuspended by pipetting, and 25µl of bead slurry per reaction was transferred to a microcentrifuge tube. Beads were equilibrated by washing using 500µl ice-cold wash/dilution buffer (in kit, 10 mM Tris/Cl pH 7.5, 150 mM NaCl, 0.05 % Nonidet™ P40 Substitute, 0.5 mM EDTA, 0.018 % sodium azide) per reaction and spun at 4^{o}C at 2500xg for 5 minutes in a microcentrifuge. This wash step was repeated twice more.

200µl lysates were diluted with 500µl wash/dilution buffer supplemented with protease inhibitor cocktail. 50µl of diluted lysate was reserved as the input fraction. The remainder was incubated with the equilibrated GFP-Trap-A beads for 1 hour at 4°C under constant mixing in a tube rotator.

After bead incubation, samples were spun at 2500xg for 5 minutes at 4°C to harvest the beads. 50µl of the supernatant was transferred to a fresh tube, labelled flow-through, the remainder was discarded. The beads were washed 3x using ice-cold wash/dilution buffer, spinning the beads at 2500xg for 5 minutes at 4°C between each wash step.

Bound protein was then eluted from the washed beads using two methods. For samples undergoing downstream mass spectrometry analysis, an acidic glycine elution buffer (200mM glycine, pH2.5) was used to elute proteins off the beads. Beads were incubated with 100µl of this buffer for 60 seconds at room temperature under constant pipetting. The bead suspension was centrifuged at 2500xg for 5 minutes, and the supernatant was collected, neutralised using 10µl 1M Tris-Base (pH 10.4) and immediately frozen at -80°C.

For elution fractions which would be subjected to western blot analysis, 100µl 2x SDS-containing western blot sample buffer was added to the beads, and beads were boiled for 5 minutes at 95°C, before centrifugation at 2500xg for 5 minutes at 4°C. Supernatant was removed and used for downstream western blot analysis.

2.15 Reverse transfection of small interfering RNA in primary fibroblasts

Small interfering RNA (siRNA) was transfected using lipofectamine RNAiMax (Thermo Fisher, Cat. No. 13778030) in Opti-MEM reduced serum media. Reverse transfection was used, where the assembled transfection complexes were added to suspended cells. For a single well of a 6-well plate, two microcentrifuge tubes were prepared, one with 250µl opti-MEM reduced serum medium and 2.5µl siRNA at 1mg/ml concentration (Dharmacon SMARTpool), and one with 250µl opt-MEM and 7.5µl RNAiMax reagent. These were incubated at room temperature for 5 minutes, then the siRNA tube was added to the RNAiMax tube. This was incubated at room temperature for 20 minutes. Meanwhile, cells were split and resuspended in complete DMEM (2ml/well). The cell suspensions were split and transfection mixtures were added dropwise to these suspension. Plates were tilted to ensure even mixing.

Cells were incubated overnight before medium was replaced by medium containing the desired treatment (detailed in figure legends). Cells were harvested 48 hours post-transfection.

RNA quantitation

2.16 RNAScope *in-situ* hybridisation

2.16.1 PROBE DESIGN

Double Z RNA probes were designed by ACDBio to hybridise with mRNAs produced from the WNT3A, WNT5A and WISP1 genes. Double Z probes have two separate elements which hybridise with different, non-overlapping sequences on the mRNA molecule. Only when two probes have hybridised with one mRNA molecule will the amplification steps work, providing stringent target specificity. All reagents and protocols were supplied by the manufacturer.

2.16.2 TISSUE PRETREATMENT AND TARGET RETRIEVAL

Formalin-fixed, paraffin-embedded lung tissue from cancer lung resections and IPF patients was sectioned and embedded on a microscope slide. All human lung experiments were approved by the Southampton and South West Hampshire and the Mid and South Buckinghamshire Local Research Ethics Committees (ref 07/H0607/73), and all subjects gave written informed consent. Tissue was deparaffinised by baking for 1 hour at 60 °C before removal of paraffin using xylene and ethanol. Tissue was treated with hydrogen peroxide for 10 minutes to block endogenous peroxidase activity, before target retrieval was performed using a proprietary reagent containing sodium citrate. This was done by submerging the slides in target retrieval reagent heated to 100 °C for 15 minutes in a steamer. The tissue was surrounded by an Immedge hydrophobic barrier for the remainder of the protocol. Target retrieval allows the RNAScope probes to access and bind to mRNA molecules inside the tissue.

2.16.3 HYBRIDISATION, AMPLIFICATION AND VISUALISATION

Probes were hybridised by dropwise addition onto pretreated tissue followed by incubation at 40 °C for 2 hours in the HybEZ II oven. Probes were washed off using wash buffer before the first amplification reagent was added, again dropwise onto the slide. 6 amplification steps were performed to add additional RNA amplification molecules to the hybridised probes. After amplification, a labelling probe conjugated to alkaline phosphatase was added, and Fast Red dye was used as the chromogenic substrate for colour development. Messenger RNA molecules can be seen under the microscope as red spots, with individual red spots corresponding to individual RNA molecules, except at very high RNA concentrations. Slides were scanned at an objective magnification of 40x by Dr David Johnston using the Olympus VS110 high throughput Virtual Microscopy System.

Figure 2.2 shows a schematic of the process of RNAScope probe hybridisation, amplification and visualisation

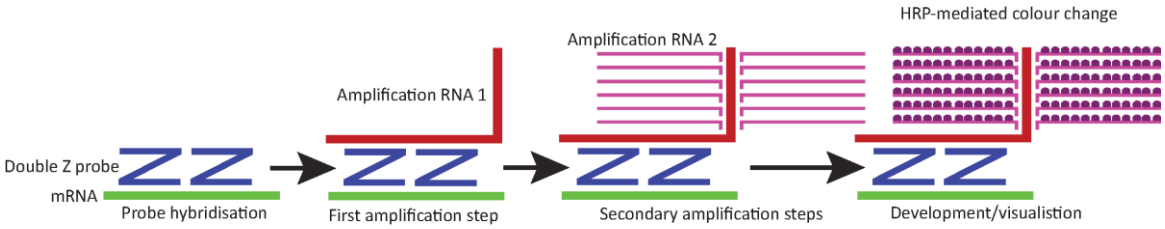


Figure 2.2: Schematic showing RNAScope probe hybridsation, amplification and visualisation on an mRNA molecule.

2.17 RNA extraction

RNA extraction was performed using Monarch total RNA miniprep kits (New England Biolabs, cat. No. T2010S) according to the manufacturer's instructions. Cells were pelleted and resuspending in 300µl of RNA lysis buffer. Lysates were added to gDNA collection tubes, then spun at 16,000 x g for 30 seconds. An equal volume of 100% ethanol was added to the flow through and mixed by pipetting. The mixture was transferred to an RNA purification column, and then spun at 16,000 x g for 30 seconds to bind RNA to the spin column. Bound RNA was treated with DNase for 15 minutes before washing with 500µl RNA priming buffer and centrifugation, followed by addition of 500µl RNA wash buffer and centrifugation. Columns were transferred to an RNase-free microcentrifuge tube, and RNA was eluted in 100µl of nuclease-free water. RNA concentrations were quantified using a NanoDrop One Spectrophotometer (Thermo Fisher, cat no. ND-ONE-W), measured using the 260nm wavelength absorbance. Purity was assessed using the ratio of 260/230nm absorbance measurements, with a higher ratio (approximately 2.0) indicative of high sample purity.

2.18 Reverse transcription

Reverse transcription to generate complementary DNA (cDNA) was done using the Precision nanoScript 2 RT kit (Primer Design, cat no. RT-Nanoscript2). This uses polyT primers to bind the polyA tail of mRNA molecules, and random nonamer primers to ensure all mRNA is amplified. To anneal primers 1µg of RNA per sample was made up to a total volume of 8µl using ddH_2O, and 1µg of each primer was added. This was incubated at 65°C for 5 minutes. To extend these primers and produce cDNA, 1µl of nanoScript 2 enzyme was added, along with 4x reaction buffer, 1µl deoxynucleotide triphosphates and 3µl ddH_2O to make the reaction volume up to 20µl. Extension

was performed in a thermocycler (BioRad T100) at 42°C for 20 minutes, followed by heat inactivation of enzyme at 75°C. cDNA samples were diluted to 5ng/µl, ready for quantitative real-time PCR (qPCR). No RNA and no reverse transcriptase enzyme negative controls were generated simultaneously.

2.19 Quantitative Real time PCR (qPCR)

Quantitative Real time PCR qPCR was performed using Primer Design primers for quantification of housekeeping genes. For fibroblast lysates, the housekeeping genes (HKG) used were UBC (ubiquitin C) and A2 (YWHAZ, phospholipase A2). 5µl Mastermix (Primer Design, cat no. PPLUS-BioRad CFX-1ml) containing nucleotides and DNA polymerase was combined with 0.5µl of primers to the housekeeping gene and 2µl ddH_2O. This was pipetted into wells of a clear 96-well qPCR plate, and 2.5µl of cDNA was added to each well. Samples were quantified in duplicate. Housekeeping genes were quantified by measuring FAM and CY5 fluorescence in a multiplex reaction. PCR reactions were carried out by incubating at 95°C for 15 seconds (strand separation), 50°C for 45 seconds (primer annealing) followed by a fluorescence reading, and 72°C for 10 seconds. This was repeated for 50 cycles in total. A BioRad CFX96 touch thermocycler was used for this. Biorad CFX software was used to calculate cycle threshold (Ct) values. These correspond to the cycle number at which a fluorescence threshold is surpassed and can be used to quantify the amount of cDNA corresponding to the gene being measured that is present in each sample using the ΔΔCt method (see below).

Thermo Fisher primers conjugated to a 6-carboxyfluorescein (FAM) fluorophore were used to quantify gene expression levels. TaqMan Fast Advance Master Mix (Thermo Fisher, cat no. A44360) was used to quantify WISP-1 variant expression. These primers are listed in .

Table 2.2. A custom assay targeting the exon 2/exon 3 WISP1 boundary was generated using the Thermo Fisher Primer Design tool found at https://www.thermofisher.com/uk/en/home/life-science/sequencing/sanger-sequencing/pre-designed-primers-pcr-sanger-sequencing.html .

Table 2.2: Assay IDs for qPCR primers targeting different WISP1 splice variants and other genes of interest (GOI)

WISP1 Variant	Assay ID
Variant 1 (full-length)	Hs00988971_m1
Variant 2	AP9HN4J (Custom Assay)
Variant 3	Hs04234731_m1
Variant 4	Hs04234732_m1
PPARG	Hs01115513_m1
LOXL2	Hs00158757_m1
PLOD2	Hs01118190_m1
RUNX2	Hs01047975_m1

Gene of interest (GOI) quantitation was done by calculating ΔCt values as follows:

$$\Delta\mathrm{Ct} = \mathrm{Ct(GOI)} - \mathrm{Ct(HKG)}$$

The ΔCt values from the control samples were then subtracted to generate ΔΔC values:

$$\Delta\Delta\mathrm{Ct} = \Delta\mathrm{Ct(sample)} - \Delta\mathrm{Ct(control)}$$

To calculate fold change in gene expression:

$$\text{Fold change in gene expression} = 2^{-\Delta\Delta\mathrm{Ct}}$$

2.20 Statistics

Before statistical testing was performed, a Kolmogorov-Smirnov test for normality was performed. If data were non-normally distributed, Wilcoxon signed-rank tests were used to compare two sample groups. Welch's T test was used to calculate the statistical significance between two groups. For samples with multiple groups, a one-way analysis of variance (ANOVA) test was performed, comparing the means of different sample groups using Dunnett's multiple comparison test. Correlation coefficients were calculated using the cor.test() function in R.

Protein quantitation

2.21 Western blotting

Total protein concentration was assessed using a Bicinchonic acid (BCA) assay (ThermoFisher, cat no. 23225). 10µl of samples was added to 200µl of BCA reagent in a 96 well plate, along with 10µl each of 8 serial 2x dilutions of bovine serum albumin (starting concentration of 2mg/ml) to produce a standard curve. Colour change was read using a MultiSkan FC plate reader at 560nm wavelength and protein concentration of samples was interpolated from the standard curve.

For SDS gel electrophoresis, samples were diluted to the same concentration as the least concentrated sample, and 6x sample buffer was added to approximately 15µg total protein. Dithiothreitol reducing agent was added to the samples at a 10% final concentration. Samples were loaded into 12.5% polyacrylamide gels (Table 2.3, Table 2.4) and run at 120V at 4 ˚C in tris-glycine running buffer (Table 2.6). Proteins were transferred from gel to nitrocellulose membranes in 4 ˚C Towbin transfer buffer (Table 2.7) and run at 300mA for 3 hours at 4 ˚C. Antibodies were diluted in blocking buffer, (Table 2.8). Membranes were washed in TBST and transfer assessed using Ponceau red protein stain, before blocking in 5% milk powder for 30 mins. Primary antibodies (Table 2.9) were diluted to an appropriate dilution (usually 1:1000) in 2% milk before incubation overnight at 4 ˚C. Membranes were washed 3x for 10 minutes in TBST before incubation with secondary antibody for 1 hour at room temperature. Secondary antibodies were either fluorescently conjugated or horseradish peroxidase (HRP) conjugated. Blots were developed using enhanced chemiluminescence reagents (HRP-conjugated antibodies) and imaged using the Amersham imager 680. Image processing was performed in imageJ.

Table 2.3: Western blot separating gel

30% (w/v) acrylamide/0.8% (w/v) bis acrylamide	4.2 ml
1.5 M Tris-HCl pH 8.8	2.5 ml
dH_2O	3.2 ml
10% (w/v) SDS	100 µl
10% APS	50 µl
TEMED	5 µl

Table 2.4: Western blot stacking gel

30% (w/v) acrylamide/0.8% (w/v) bis acrylamide	0.5 ml
0.5 M Tris-HCl pH 6.8	1.25 ml
dH_20	3.05 ml
10% (w/v) SDS	50 ml
10% APS	25 ml
TEMED	10 ml

Table 2.5: 5x sample buffer

0.3125M Tris-HCl pH6.8	10.41ml;1.5M
50% glycerol	25ml
25% 2-mercaptoethanol	12.5ml
10% SDS	5g
0.01% bromophenol blue	5mg
dH_20	Make up to 50 ml

Table 2.6: 10x Tris-glycine running buffer

0.025M Tris	15.5.g
0.192M glycine	720g
0.1% (w/v) SDS	250ml 20%SDS
pH 8.3 with HCl	
dH_20	Make up to 5 litres

Table 2.7: 10x Towbin Transfer buffer recipe. Dilute 1:10 for a 1x working concentration, 700ml dH_2O, 200ml Methanol and 100ml 10x buffer.

0.025M Tris	151.5g
0.192M glycine	720g
dH_20	Make up to 5 litres. Do not adjust

Table 2.8: 10x Tris buffered saline (TBS). Dilute 1:10 in dH_2O for a 1x working concentration, add 1ml Tween-20 per litre

Tris	24.2g
NaCl	80g

Blocking buffer – 5% powdered milk (Marvel) in TBST

Table 2.9: Antibodies used for western blotting. (2^o) labelled antibodies are HRP-conjugated secondary antibodies. Antibodies were diluted in blocking buffer.

Target	Concentration	Manufacturer	Cat no.
WISP-1	0.1µg/ml	R&D	AF1627
FGF2	0.1µg/ml	Abcam	ab208687
PHB	0.1µg/ml	Abcam	ab75766
GFP	0.1µg/ml	Abcam	ab5450
VDAC1	0.1µg/ml	Abcam	ab15895
β-actin	5ng/ml	Cell signalling	5125S
Rabbit IgG (2^o)	1:5000	Cell signalling	7074S
Mouse IgG (2^o)	1:5000	Cell signalling	7076S
Goat IgG (2^o)	1:5000	Invitrogen	PA1-28664

2.22 Immunohistochemistry

Immunhistochemistry and tissue staining was performed by Dr Christopher Brereton.

FFPE tissue sections were deparaffinised by incubation with Clearene solvent (Leica biosystems, Cat no. 3803600) for 2x 10 minutes. They were then rehydrated using gradated alcohols to a final incubation with deionised water.

Antigen retrieval was performed using 0.01M citrate buffer pH 6 in a microwave at 50% power for 20 minutes. Endogenous peroxidase activity was inhibited by incubation in 3% hydrogen peroxide in deionised water for 15 minutes, followed by rinsing with PBS.

Blocking of non-specific antigens was performed by incubating slides in goat serum (Dako, cat no. X090710-8) at a concentration of 0.015% in TBST, as a goat-derived secondary antibody was used. This was tapped off before incubation overnight at 4^oC with an anti-CA9 rabbit primary antibody (Novus Bio, cat no. NB100-417) diluted 1:500 in blocking solution. Non-specific staining was excluded using rabbit IgG at the same concentration as the primary antibody as an isotype control. Primary antibody was tapped off before rinsing slides with TBS for 3x 5 minutes. Biotinylated goat anti-rabbit secondary (Dako, cat no. E0432) at a 1:500 dilution in blocking buffer

was applied to slides (150µl/slide) for 20 minutes, before rinsing with TBS 3x 5 minutes. Slides were incubated at room temperature for 30 minutes in 150µl of Avidin Biotin Complex reagent (Vectastain, Elite ABC Kit, Cat no. PK6100) before rinsing 3x 5 minutes in TBS. Slides were incubated for 5 minutes with 150µl/slide 3,3'Diaminobenzidine chromogenic substrate (Dako, cat no. K3468) before rinsing with TBS. Slides were washed in running water for 2-3 minutes before section were counterstained with Mayer's Haematoxylin (Sigma, cat no. MHS32) for 1 minute, rinsed for 2 minutes in running water and dehydrated using gradated alcohols to clearene. Pertex mounting medium was used to mount a coverslip, before imaging on the Olympus VS110 slide scanner.

Haematoxylin and eosin (H&E) staining allows differentiation of cellular structures by staining nuclei blue and cytoplasm and ECM pink. For FFPE tissue staining, sections were deparaffinised as above, before incubation in Mayer's Haematoxylin for 5 minutes, running tap water for 5 minutes, and eosin (Sigma, cat no E4009) for 5 minutes. Sections were then washed 3x in distilled water, 3x in 100% ethanol, before dehydration with ethanol and clearene. Sections were mounted using pertex.

2.23 Mass spectrometry

2.23.1 PROTEIN QUANTITATION AND REDUCTION/ALKYLATION/DIGESTION

Protein samples were quantified using a Direct Detect infrared protein spectrometer (Merck), and the same amount of protein was prepared for each sample for downstream mass spectrometry. Approximately 5µg of elution fraction was prepared. Urea/thiourea/HEPES (pH 7.5) solution was used to denature proteins, and proteins were reduced by incubating denatured protein solution (100µl) with 0.1mM DTT for 1 hour at room temperature. Reduced -SH groups were alkylated using 55µM iodoacetamide for 45 mins in the dark at room temperature to increase peptide yield. Samples were diluted in 400µl 20mM ammonium bicarbonate to stabilise pH, and proteins were digested using 2µg trypsin per sample (Promega, cat. No. V5111) overnight at 37^{o}C.

2.23.2 C18 CLEAN UP

Samples containing tryptic peptides were acidified using 100% trifluoroacetic acid, added in 1µl amounts until pH as measured using litmus paper was below 3.0. Samples were lyophilised at 50^{o}C in a SpeedVac until samples were dry (Thermo Fisher). Samples were desalted and all detergents and buffers removed using C18 96-well clean-up plates (Waters, Cat. No. 186008304). Firstly 100µl methanol was added to each well to be used to wet the silica in the clean-up membranes. This was drawn through the plate under vacuum, followed by 200µl of 80% acetonitrile, 0.5% acetic acid solution to permeabilise the membrane. Membranes were washed using 200µl of 0.5% acetic acid solution. Lyophilised samples were resuspended in 100µl 0.1%

acetic acid and loaded onto clean-up plate (one well per sample). Liquid was drawn through under vacuum, and peptides bound to the plate membrane were washed twice with 200µl 0.5% acetic acid. Samples were eluted using 150µl 80% acetonitrile solution and transferred to microcentrifuge tubes. Cleaned-up samples were again lyophilised until dry using the SpeedVac. Lyophilised samples were resuspended in 10 µl 0.1% formic acid and stored at 4^{o}C until ready to run.

2.23.3 MASS SPECTROMETRY

Mass spectrometry was performed by Dr Benjamin Nicholas. Samples were subjected to high pressure liquid chromatography to separate tryptic peptides, and mass spectra were obtained using an Orbitrap mass spectrometer (Thermo Fisher). Proteome Discoverer software was used both to assign peptide identities to mass spectra and as a peptide search engine to assign peptides to specific proteins. Protein and peptide data were exported as a .csv file, and protein lists were extracted using R.

2.24 Protein BLAST

Amino acid sequences from two proteins to be compared were entered into the NCBI's protein basic local alignment search tool (BLAST) and sequence similarity compared using the protein BLAST algorithm.

2.25 Protein interaction WISP-1 ELISA

In order to generate binding curves for the interaction of WISP-1 with basic fibroblast growth factor (FGF2), the methodology used in Nishida et al. to quantify the binding of connective tissue growth factor (CTGF) to FGF2 was adapted to WISP-1(183). Components from a WISP-1 ELISA kit (R&D, cat no. DY1627) was used to quantify WISP-1 binding to FGF2. To perform the assay, 8 serial dilutions of recombinant FGF2 (abcam, cat no. ab9596) were used to coat the wells of an ELISA plate (in reagents kit, R&D, cat no. DY008) overnight at room temperature, halving the concentration each time, from 6µg/ml to 0.047µg/ml (353nM – 2.75nM). Blank wells without any FGF2 were treated similarly. Wells were washed 3x using ELISA wash buffer (0.05% Tween-20 in PBS). Remaining wash buffer was removed by blotting on paper towels. Plates were blocked for 1 hour using 300µl plate coating buffer (1% BSA in PBS). Plates were washed 3x as above. 100µl of recombinant WISP-1 (R&D, cat no. 1627-WS-050) at a concentration of 1µg/ml was added to wells, and incubated at room temperature for 2 hours, followed by 3x plate washes. 100µl of detection antibody (biotinylated ?species anti WISP-1) at a concentration of 20ng/ml was added to wells and incubated at room temperature for 2 hours, followed by 3x plate washes. 100µl of

streptavidin-HRP solution was added to wells and incubated for 20 minutes at room temperature in the dark before 3x plate washes. 100µl of substrate solution, a 1:1 mixture of H_2O_2 and tetramethylbenzene was added to wells and incubated in the dark until visible colour change was observed. 50µl of stop solution (1M H_2SO_4) was added to wells to stop colour change reaction.

Absorbance of each wells was measured using a MultiSkan FC plate reader (Thermo Fisher, cat no. 51119000) at 450nm wavelength. Wavelength readings at 570nm were subtracted from 450nm measurements to correct for non-specific emissions. Measurements were blank corrected, and absorbance measurements from wells treated with WISP-1, but only coated with 1% BSA solution in PBS, were subtracted to remove the effect of non-specific binding.

3 Analysis of IPF transcriptomic datasets

3.1 Chapter introduction and aims

In recent years, several transcriptomic datasets have been published which have shed light on the transcriptional landscape of IPF. Studies have identified IPF-specific gene signatures (184), differential splicing events in IPF (185), and candidates for IPF-specific genes, including matrix metalloproteinases(186).

The purpose of this chapter is twofold – firstly, to identify the tissue areas and cell types associated with WISP-1 expression in IPF tissue in several different datasets. This will both allow the cells which express WISP1 to be identified, so an appropriate model can be used for downstream affinity purification-mass spectrometry analysis, and also to characterise these cells and tissue types, identifying the gene expression environment in which WISP-1 can be found

The second aim of this chapter is to more broadly characterise the cell and tissue types present in IPF – which cell types are abundant, how they are distributed within IFP tissue, and what their roles in pathology might be. Much of this analysis has been restricted to the various fibroblast types found in IPF due to their status as the leading edge of IPF pathogenesis, but some effort has been devoted to identifying cell types present in IPF alveolar septae as well.

To do this, several different datasets have been used. The first part of this chapter is a brief meta-analysis comparing differentially expressed genes in IPF vs control lung samples in two separate bulk-tissue transcriptomic datasets, identifying genes and gene ontology terms common across multiple datasets.

3.2 Meta analysis comparing bulk IPF transcriptomic datasets

In order for a comprehensive study of the gene expression changes undergone by the lung in IPF, it was first necessary to identify common gene expression signatures associated with IPF when compared to normal lung tissue. This was done by identifying which genes are differentially expressed in both datasets. This has added value in that it can also be used to compare to the differentially expressed genes and gene ontology terms associated with the laser capture microdissection RNAseq analysed below. From the perspective of WISP1 expression, it was previously observed in Selman et al (2006) that WISP1 is upregulated in IPF compared to hypersensitivity pneumonitis(72). This dataset was not reanalysed as this dataset was not

comparing IPF lung with normal lung, but instead investigated changes in IPF lung versus hypersensitivity pneumonitis. While it identified WISP1 expression as being upregulated in IPF versus HP, it does not identify gene expression changes in IPF compared to normal lung tissue. Identifying these changes is key to understanding the pathogenesis of IPF, so datasets which compared normal lung tissue with IPF lung tissue were chosen instead.

Datasets containing only IPF and control lung, and only bulk tissue, were selected, with other datasets which contained other ILD samples, such as GSE32537, being discounted.

In order to identify common differentially expressed genes between IPF and control lungs, two datasets were used. These were both bulk transcriptomic datasets taken from the Gene Expression Omnibus (GEO), one using RNA sequencing and one using microarray data. The datasets were taken from the GEO repositories GSE24206 (184), and GSE52463 (185). They are microarray and RNA-seq datasets respectively. The goals of these two studies was not simply to identify differentially expressed genes in IPF – Melzter et al. was designed to quantify differences between different lung lobes, and differences in gene expression between explant and biopsy samples in IPF. Nance et al. used RNA sequencing with a high sequencing depth on a smaller number of IPF samples to identify differential splicing events in IPF. Details on the two studies are presented below.

Table 3.1: Details of two transcriptomic studies looking at IPF vs control lung.

Study	**Number of samples**	**Sequencing Methodology**
Meltzer et al. 2011	17 samples from 11 IPF patients, of which 6 were biopsies, 5 explants. 6 controls from donor lungs.	Microarray – Affymetrix Human Genome U133 Plus 2.0
Nance et al. 2014	8 IPF lung samples, 7 Healthy controls	RNA-sequencing – Illumina HiSeq 2000

The processed data (raw counts for RNA-seq data, Log2-transformed microarray intensities for microarray data) were downloaded from their respective GEO repositories and differential expression performed on them. The R package Limma was used for differential expression of microarray data, and EdgeR was used for differential expression of RNA-seq data.

3.3 Upregulated genes in bulk transcriptomic datasets include extracellular matrix components and matrix metalloproteinases

Table 3.2 shows the top upregulated genes identified in Meltzer et al. Several genes previously associated with IPF, including MMP7 are very strongly upregulated, as well as genes such as COL1A1, COMP and ASPN which code for extracellular matrix components, and SFRP2, a secreted frizzled receptor protein gene associated with antagonism of Wnt signalling. The upregulated genes reflect the increased extracellular matrix production which is a hallmark of IPF.

Table 3.2: Top upregulated genes in IPF compared to control tissue in Meltzer et al. (GSE24206 data). Comparison made using Limma R package, genes are ordered by Log2 fold-change values (LogFC). T is t statistic, adjusted p values were calculated using Benjamini-Hochberg multiple test correction.

Gene Symbol	logFC	Average Expression	t	P.Value	adj.P.Val
MMP7	3.40	9.46	5.09	3.25E-05	0.001698
CXCL14	3.29	8.08	5.58	9.65E-06	0.000861
SFRP2	3.26	10.46	5.15	2.80E-05	0.001576
SERPIND1	3.25	6.94	6.62	7.41E-07	0.000197
COL1A1	2.98	7.70	7.25	1.68E-07	9.09E-05
MUC5B	2.81	8.31	3.77	0.000949	0.013922
JUP	2.75	7.37	4.80	6.90E-05	0.002729
ASPN	2.60	10.13	8.66	7.35E-09	1.29E-05
HS6ST2	2.56	6.67	5.46	1.30E-05	0.000985
GABBR1	2.51	7.68	7.28	1.59E-07	8.98E-05
COL15A1	2.51	9.56	7.88	4.07E-08	3.59E-05
S100A2	2.47	8.75	2.95	0.007024	0.049538
LPPR4	2.47	6.86	8.00	3.09E-08	3.01E-05
COMP	2.41	7.08	5.83	5.18E-06	0.000605
CRIP1	2.40	10.68	13.35	1.29E-12	3.52E-08
EPHA3	2.38	6.26	6.77	5.29E-07	0.000164

PROM2	2.36	7.36	3.86	0.000757	0.012051
CTHRC1	2.33	9.00	5.41	1.46E-05	0.001063
FAM81B	2.31	7.05	3.17	0.004136	0.035466
PSD3	2.27	7.51	12.21	8.41E-12	1.53E-07
SPATA18	2.26	8.41	3.77	0.000931	0.013753
CCDC113	2.23	7.12	3.05	0.005462	0.042333
CXCL13	2.18	6.84	3.16	0.004213	0.035867
LOC102725271	2.17	9.12	3.02	0.005943	0.044666
FNDC1	2.16	5.65	11.84	1.61E-11	1.76E-07

Table 3.3 shows the top upregulated genes in IPF samples compared to control in the Nance et al. RNA-seq dataset. Certain gene classes such as SERPINs and matrix metalloproteinases are present in the top genes from both datasets, but this data has a strong immune signature, with several immunoglobulin genes upregulated which are not present in the previous dataset.

Table 3.3: Top upregulated genes in IPF compared to control lung in Nance et al. (GSE52463 data). Comparison made using EdgeR R package. Genes are ordered by log2 fold change (logFC). False discovery rate p values calculated using Benjamini-Hochberg multiple test correction.

Gene Symbol	LogFC	LogCPM	F	PValue	FDR
IGHG4	6.83	7.27	66.22	3.43E-08	0.000196
LGALS7	6.18	-0.68	18.93	0.000447	0.042214
MUC5B-AS1	5.20	3.89	28.58	2.05E-05	0.007361
IGHV1-12	5.13	0.46	19.53	0.000202	0.027423
IGHV1-69	5.00	6.02	48.08	4.79E-07	0.000912
MMP1	4.84	5.78	25.20	4.56E-05	1.16E-02
IGLV2-23	4.72	6.33	42.14	1.32E-06	1.57E-03
IGHG1	4.62	10.89	56.23	1.36E-07	4.65E-04
MUC5B	4.58	8.32	19.97	0.000178	2.57E-02
IGHV1-3	4.56	4.66	25.84	3.90E-05	1.05E-02
IGHV1-2	4.39	4.82	28.92	1.89E-05	6.93E-03

IGHGP	4.37	3.96	39.19	2.28E-06	1.83E-03
COPDA1	4.34	5.93	77.75	8.40E-09	1.24E-04
SERPINB2	4.27	3.85	22.68	8.60E-05	1.68E-02
IGHG2	4.19	7.76	78.38	7.82E-09	1.24E-04
IGKV1-17	4.17	4.21	37.82	2.95E-06	2.23E-03
KRT14	4.17	2.51	22.24	9.65E-05	1.79E-02
IGKV1D-33	4.13	1.03	29.98	1.49E-05	6.09E-03
IGKV1-39	4.12	2.81	52.29	2.45E-07	6.25E-04
IGKJ3	4.11	-0.62	28.99	1.86E-05	6.93E-03
IGKV1-12	4.10	4.77	31.71	1.02E-05	4.95E-03
IGHJ3	4.09	1.47	36.37	3.92E-06	2.69E-03
IGLC3	4.09	7.48	59.44	8.56E-08	4.00E-04
CYP24A1	4.06	4.03	20.16	0.000169	2.49E-02

3.4 There is low overlap between differentially expressed genes in bulk IPF datasets

Due to the differing purposes and methodologies of the two datasets, some discrepancy between them is not unprecedented. However, Figure 3.1 shows the relatively small overlap between significantly differentially expressed genes present in both datasets. There is a common 79-gene signature which is significantly differentially expressed in both datasets, with the majority of differentially expressed genes being exclusive to the Meltzer et al. microarray dataset. Methodological differences, such as consistency of microarray results and the increased sample sizes of the Meltzer et al. dataset account for some of this difference, but the majority (176) of differentially expressed genes in the Nance et al. RNA-seq dataset are not present in the Meltzer et al. list.

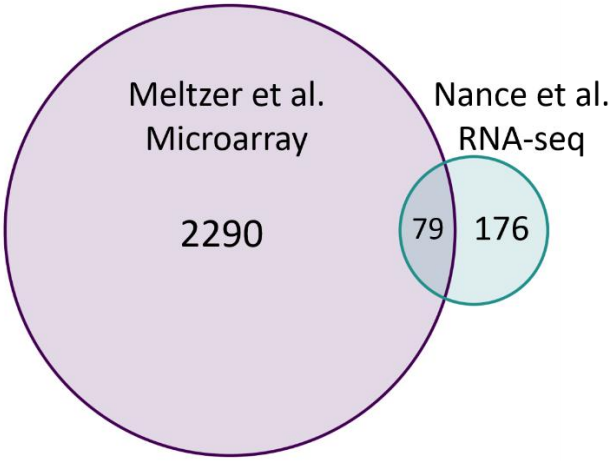


Figure 3.1: Venn diagram showing overlap between upregulated genes in IPF samples in two different transcriptomic datasets. Upregulated genes were defined by FDR-adjusted P-value of < 0.05 following differential expression analysis using Limma (microarray data) or EdgeR (RNA-seq data).

3.5 Common upregulated genes show a recognisable IPF-specific signature

Table 3.4 shows the top 25 common upregulated genes from both datasets, ordered by mean log2 fold change. MUC5B, is at the top of the list. Polymorphisms in the MUC5B gene are a strong predictor of IPF development, so its differential expression is intriguing. Mucociliary dysfunction is a hallmark of IPF, associated with MUC5B dysregulation(187). MMP7, a gene previously associated with IPF is highly upregulated, and the fibrotic nature of the disease is reflected by the number of genes associated with extracellular matrix development, including COMP (Cartilage Oligomeric Matrix Protein) and ASPN (Asporin), a cartilage-associated protein(76,188). SFRP2, a Wnt-induced Wnt signalling antagonist, is also strongly upregulated, affirming the previously identified Wnt signalling signature in IPF. Notably, however, WISP1 is not differentially expressed in either dataset, although its expression was detected in both.

Table 3.4 Top common upregulated genes in IPF samples compared to control ordered by mean log2 fold change in two different transcriptomic datasets.

	Meltzer et al.		Nance et al.		
Gene Symbol	**LogFC**	**FDR**	**LogFC**	**FDR**	**Mean logFC**
MUC5B	2.81	0.01	4.58	0.03	3.69
CXCL14	3.29	0.00	4.01	0.00	3.65
MMP7	3.40	0.00	3.50	0.01	3.45
SFRP2	3.26	0.00	3.38	0.00	3.32
CHIT1	2.12	0.04	3.96	0.02	3.04
COMP	2.41	0.00	3.44	0.00	2.92
S100A2	2.47	0.05	3.29	0.04	2.88
FNDC1	2.16	0.00	3.16	0.00	2.66
PROM2	2.36	0.01	2.21	0.01	2.28
ASPN	2.60	0.00	1.83	0.01	2.22
DIO2	1.61	0.00	2.75	0.00	2.18
COL14A1	2.13	0.00	2.17	0.00	2.15
EPHA3	2.38	0.00	1.89	0.01	2.13
SLC28A3	1.92	0.01	2.15	0.00	2.03
ALDH1A3	1.58	0.02	2.38	0.00	1.98
CRABP2	0.88	0.01	3.04	0.00	1.96
PDLIM4	1.68	0.00	2.15	0.00	1.92
CDH3	1.86	0.00	1.96	0.02	1.91
MEOX1	1.55	0.01	2.26	0.02	1.90
EYA2	1.52	0.04	2.12	0.02	1.82
LRRN1	1.72	0.00	1.79	0.00	1.75
SULF1	1.70	0.00	1.75	0.00	1.73
CBLN4	1.42	0.01	1.98	0.01	1.70
PCDH7	1.93	0.00	1.39	0.01	1.66

Table 3.5 shows the top gene ontology terms associated with these 76 genes, comparing these genes with a background list containing only genes whose symbols are present in both datasets. The pro-fibrotic nature of IPF is very clearly demonstrated, with multiple GO terms associated

with extracellular matrix development and collagen development, reflecting the dysregulated collagenous extracellular matrix deposition which is a hallmark of the disease.

Also notable are the terms associated with signalling pathways known to be important for IPF pathogenesis. Genes for TGF-β signalling activation and production are upregulated in IPF in both datasets, highlighting the important role TGF-β is known to have in IPF. The only Wnt-related term is negative regulation of non-canonical Wnt signalling. However, Canonical Wnt signalling has been shown to both negatively regulate non-canonical Wnt signalling and upregulate Wnt antagonists such as the Dickkopf (DKK) proteins and SFRPs, so this may be reflective of homeostatic Wnt-controlling responses to high Wnt signalling(189,190).

Finally, there are several terms associated with chondrocyte development, as well as skeletal system development. These are potentially interesting as recent research into collagen stiffness in IPF has identified bone-type collagen crosslinking as being important for the stiffened , inflexible collagen typical of the disease(65).

Table 3.5: Top gene ontology terms associated with common upregulated genes in two IPF transcriptomic databases. Gene ontologies were calculated using GOrilla using the list of common upregulated gene symbols as the target set, and all gene symbols present in both gene sets as the background set. Only cellular process gene ontology terms were queried.

Gene Ontology Description	**P-value**	**FDR q-value**
Extracellular matrix organization	4.69E-08	7.17E-04
Extracellular structure organization	2.14E-07	1.64E-03
Chondrocyte development	3.87E-06	1.97E-02
Oesophagus smooth muscle contraction	2.70E-05	1.03E-01
Developmental process	3.59E-05	1.10E-01
Collagen fibril organization	6.65E-05	1.70E-01
Negative regulation of cellular response to growth factor stimulus	7.71E-05	1.68E-01
Sclerotome development	8.06E-05	1.54E-01
Anatomical structure development	1.21E-04	2.06E-01
Smooth muscle contraction	1.33E-04	2.03E-01
Skeletal system development	1.37E-04	1.91E-01
Cell adhesion	1.62E-04	2.07E-01
Biological adhesion	1.73E-04	2.03E-01
Animal organ morphogenesis	1.80E-04	1.96E-01
Saliva secretion	2.67E-04	2.72E-01
Regulation of cellular response to growth factor stimulus	2.83E-04	2.70E-01
Cartilage development	3.46E-04	3.12E-01
Gastro-intestinal system smooth muscle contraction	3.99E-04	3.39E-01
Regulation of tooth mineralization	3.99E-04	3.21E-01
Negative regulation of non-canonical Wnt signalling pathway	3.99E-04	3.05E-01
Regulation of transforming growth factor beta activation	3.99E-04	2.91E-01
Regulation of transforming growth factor beta production	5.56E-04	3.87E-01
Bone morphogenesis	5.56E-04	3.70E-01
Glomerular filtration	5.57E-04	3.55E-01
Renal filtration	7.40E-04	4.53E-01
Chondrocyte proliferation	7.40E-04	4.35E-01
Glial cell-derived neurotrophic factor receptor signalling pathway	7.40E-04	4.19E-01

Although IPF-specific gene signatures can be identified from these datasets, their nature – using bulk tissue as a source of RNA - means that any heterogeneity within IPF tissue beyond large differences in the tissue location such as lung lobe cannot be investigated. This is problematic as IPF is known to be a disease of significant heterogeneity at the microscopic level, with areas of normal-appearing lung tissue adjacent to areas of fibrosis, as well as cellular aggregates like fibroblast foci. This makes investigation of specific genes such as WISP1, which may have low expression levels in most areas of lung tissue, hard to investigate. The use of laser-capture microdissection RNA-seq data and single-cell RNA-seq data, below, seek to address this issue.

3.6 Laser Capture microdissection data

Laser capture microdissection microscopy was used to produce a transcriptomic dataset of precisely excised areas of normal and IPF lungs. Alveolar septae from Formalin-fixed, paraffin-embedded control lungs and IPF biopsies, and fibroblast foci from IPF lungs were sampled. Multiple areas of the tissue type in question were excised, and these were pooled to give sufficient mRNA for sequencing. Tissue procurement and laser capture microdissection were performed at the University of Southampton by Dr Mark Jones. mRNA processing and sequencing were performed at the University of Yale. After sequencing and read assignment, fragments per kilobase per million mapped reads (FPKM) values were calculated from raw counts data, normalising both for sequencing depth and gene length. 10 control and 10 IPF tissue samples were used, with alveolar septae and fibroblast foci being dissected from the same IPF tissue samples, leading to 30 samples in total.

Sequencing was performed using an Ion Torrent Proton sequencer. Approximately 20 million single-end reads of roughly 100bp in length were produced per sample. Post processing was performed using Cufflinks, and reads were mapped to the University of California Santa Cruz hg19 human genome before calculation of fragments per kilobase of exon per million mapped reads (FPKM) values. RNA integrity (RIN) numbers were also measured to assess RNA quality prior to sequencing.

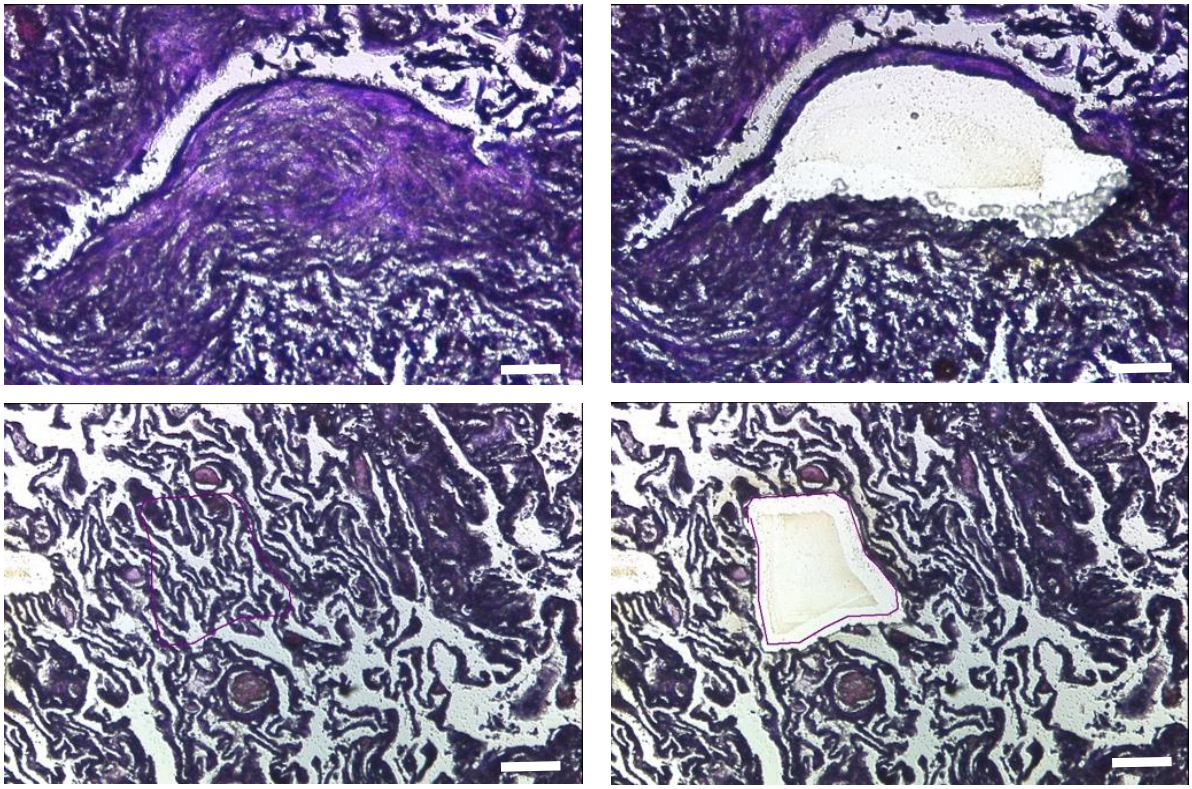


Figure 3.2: Before (left) and after (right) micrographs showing laser capture microdissection (LCMD) of IPF lung tissue. Top images show a fibroblast focus before and after excision by laser capture microdissection, bottom images show an area of IPF alveolar septae before and after LCMD. Images courtesy of Dr Mark Jones, University Hospital Southampton.

3.7 Laser-capture microdissection data segregates according to sample type on a principal component analysis plot

Principal component analysis (PCA) was performed on these data to identify the axes of greatest variation in the dataset. Plotting the first two principal components of this analysis showed a strong segregation of the different sample types, with IPF alveolar septae, control alveolar septae and fibroblast foci clustering separately. Notably, the fibroblast foci show a stronger segregation from the two alveolar clusters than they do to each other, suggesting that a major source of variation in this dataset is the difference between the fibroblast foci and alveolar samples. Two of the IPF alveolar septae samples also showed significant deviation from the others on the PCA plot, suggesting that these samples differ in some ways from the other samples in this group.

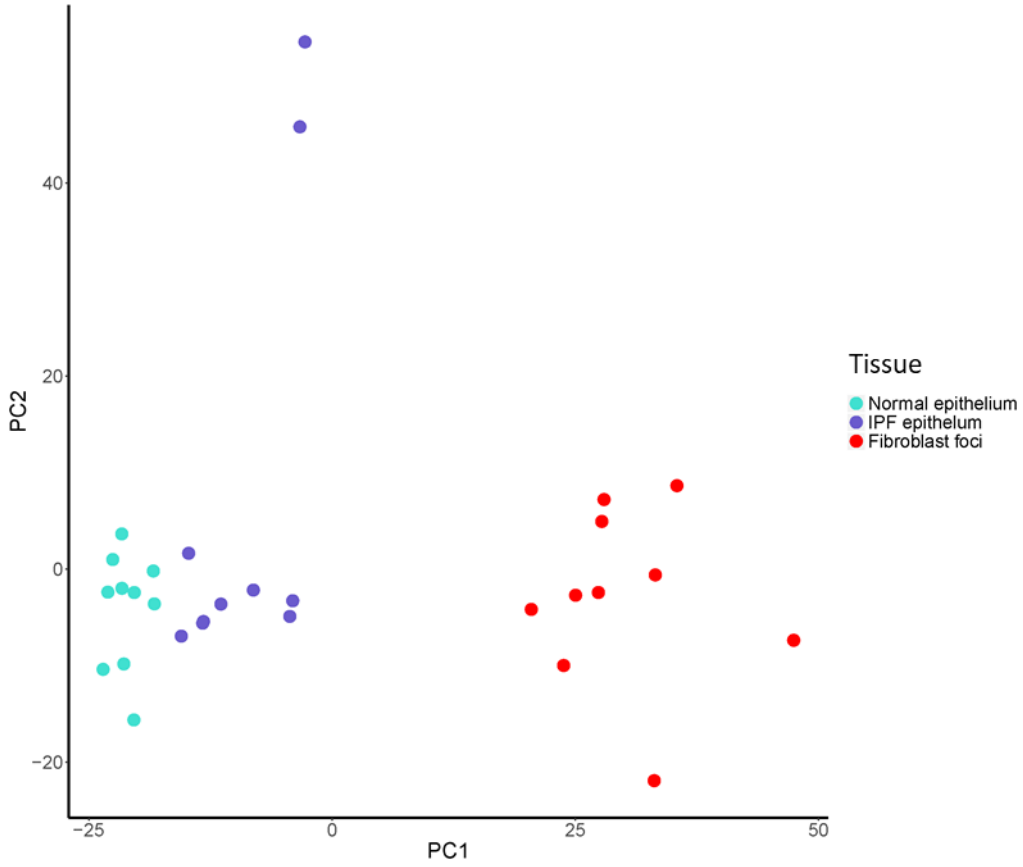


Figure 3.3: Principal component analysis (PCA) plot showing variance of different laser-capture microdissection samples. PC1 and PC2 are the first and second principal components, respectively. Point colour represents the tissue type.

3.8 Comparison of alveolar samples shows IPF-specific gene signatures

Initial differential expression analysis of this dataset focussed on comparing IPF alveolar septae with control alveolar septae, as this allowed comparison of analogous tissue types, without the confounding influence of fibroblast gene signatures from the fibroblast foci. Gene signatures enriched in IPF alveolar septae showed an enrichment in functions associated with extracellular matrix development. This is supported by two of the top differentially expressed genes associated with IPF alveolar development being a collagen gene, COL3A1, and ASPN (Asporin) a protein associated with cartilage development. The third most upregulated protein, SFRP4, is a Wnt antagonist which has previous been associated with modulation of Wnt signalling, and regulation of apoptosis. It may have importance in regulating cellular proliferation in certain types of cancer.

Table 3.6: Top ranked upregulated genes and gene descriptions in IPF alveolar septae LCMD data compared to control alveolar septae. Gene set enrichment analysis was performed using the Broad institute's GSEA software. Signal2Noise metric was used for calculating gene ranking score, with the default 1000 permutations.

Gene	Description	GSEA Ranking Score
THY1	Thy-1 cell surface antigen	2.46
FSTL1	follistatin like 1	2.08
COL3A1	collagen type III alpha 1 chain	1.99
SFRP4	secreted frizzled related protein 4	1.90
CFH	complement factor H	1.90
ASPN	asporin	1.80
KCTD12	potassium channel tetramerization domain containing 12	1.75
CST2	cystatin SA	1.72
GAL3ST4	galactose-3-O-sulfotransferase 4	1.70
RARRES1	retinoic acid receptor responder 1	1.60
CDR1	cerebellar degeneration related protein 1	1.59
COL17A1	collagen type XVII alpha 1 chain	1.57
FAM150A	ALK and LTK ligand 1	1.56
F2RL2	coagulation factor II thrombin receptor like 2	1.55
SYNM	synemin	1.54
ATP5EP2	ATP Synthase F1 Subunit Epsilon Pseudogene 2	1.53
PTENP1	phosphatase and tensin homolog pseudogene 1	1.53
MEIS3P1	Meis homeobox 3 pseudogene 1	1.49
IGFBP7	insulin like growth factor binding protein 7	1.48
THSD4	thrombospondin type 1 domain containing 4	1.45
C1orf54	chromosome 1 open reading frame 54	1.43
ELN	elastin	1.42
FAT1	FAT atypical cadherin 1	1.42
ITGAV	integrin subunit alpha V	1.41
TSHZ2	teashirt zinc finger homeobox 2	1.38
PLXNB3	plexin B3	1.37

3.9 Gene set enrichment analysis (GSEA) showed upregulation of genes associated with fibrosis and Wnt signalling

Differential expression analysis using Limma was used for an initial assessment of differences between samples sets in the laser capture dataset. This allowed for a robust comparison of the normalised FPKM values provided. However, in order to gain a more representative idea of the gene sets and gene ontology terms associated with different parts of this dataset, gene set enrichment analysis (GSEA) was used to identify enriched molecular phenotypes. This utilises a ranked list of genes based on differential expression between predefined phenotypes – in this case control alveolar septae and IPF alveolar septae. GSEA takes as input a predefined list of genes associated with a particular process – in this case, gene ontologies and Hallmark datasets were used. As GSEA takes into account all genes associated with a particular phenotype, it allows a more holistic and quantitative way of identifying enriched phenotypes than classical differential expression – enrichment analysis.

Comparison of the two alveolar septae groups using GSEA showed an upregulation of gene ontology terms associated with fibrosis (Table 3.7). This is in broad agreement with the previous differential expression analysis but provides an extra level of detail. Many of these gene ontology terms are associated with extracellular matrix development, a strong indicator of the increased deposition of extracellular matrix seen in IPF tissue. Notable genes enriched in this phenotype include multiple collagen genes, as well as lysyl oxidase, an important collagen crosslinking enzyme. Stiffer extracellular matrix due to increased bone-type collagen cross linking has been previously identified as a hallmark of IPF pathogenesis, and bone development is an important phenotype associated with WISP-1(65).

Notably, enriched GO terms include Wnt-activated receptor activity that are linked to developmental and wound healing responses. Expression of several Wnt receptor genes, including several different frizzled receptors, is significantly upregulated. This fits in well with prior research that has identified and increase in Wnt signalling as being important in IPF pathogenesis. It also provides a potential mechanism for driving WISP-1 expression, as current evidence suggests its expression is Wnt-induced. However, investigation of WISP1 gene expression in the laser-capture microdissection dataset shows that it is not expressed in alveolar septae in either control or IPF tissue.

Table 3.7: Top enriched gene ontology terms in IPF alveolar septae LCMD data compared to control alveolar septae. Gene sets are ranked by normalised enrichment score, calculated using the GSEA software based on a ranked list of genes. Gene ranking was calculated using the Signal2Noise metric.

Gene Ontology Description	NOM p-val	FDR q-val
Collagen fibril organization	0.01	1.00
Wnt protein binding	0.01	1.00
Extracellular matrix structural constituent	0.03	0.94
Wnt activated receptor activity	0.01	0.74
Mesenchyme morphogenesis	0.01	0.79
Brain morphogenesis	0.01	0.72
Renal system vasculature development	0.02	0.91
Collagen trimer	0.04	0.81
Detection of temperature stimulus	0.00	0.91
Complex of collagen trimers	0.03	0.85
Endocardial cushion morphogenesis	0.01	0.78
Multicellular organismal macromolecule metabolic process	0.04	0.85
Hydrogen peroxide catabolic process	0.01	0.80
Dentate gyrus development	0.01	0.76
Regulation of hair follicle development	0.00	0.73
Extracellular structure organization	0.05	0.71
Multicellular organism metabolic process	0.03	0.73
Aorta morphogenesis	0.02	0.73
Extracellular matrix component	0.05	0.72
Fibril organization	0.03	0.70
Enteroendocrine cell differentiation	0.04	0.70
Cochlea morphogenesis	0.01	0.68
Embryonic eye morphogenesis	0.01	0.70
Extracellular matrix disassembly	0.03	0.68
Proteoglycan binding	0.02	0.66
Cochlea development	0.01	0.76
Fibronectin binding	0.05	0.75
Negative regulation of stem cell proliferation	0.02	0.81
Wnt signalling pathway calcium modulating pathway	0.01	0.80

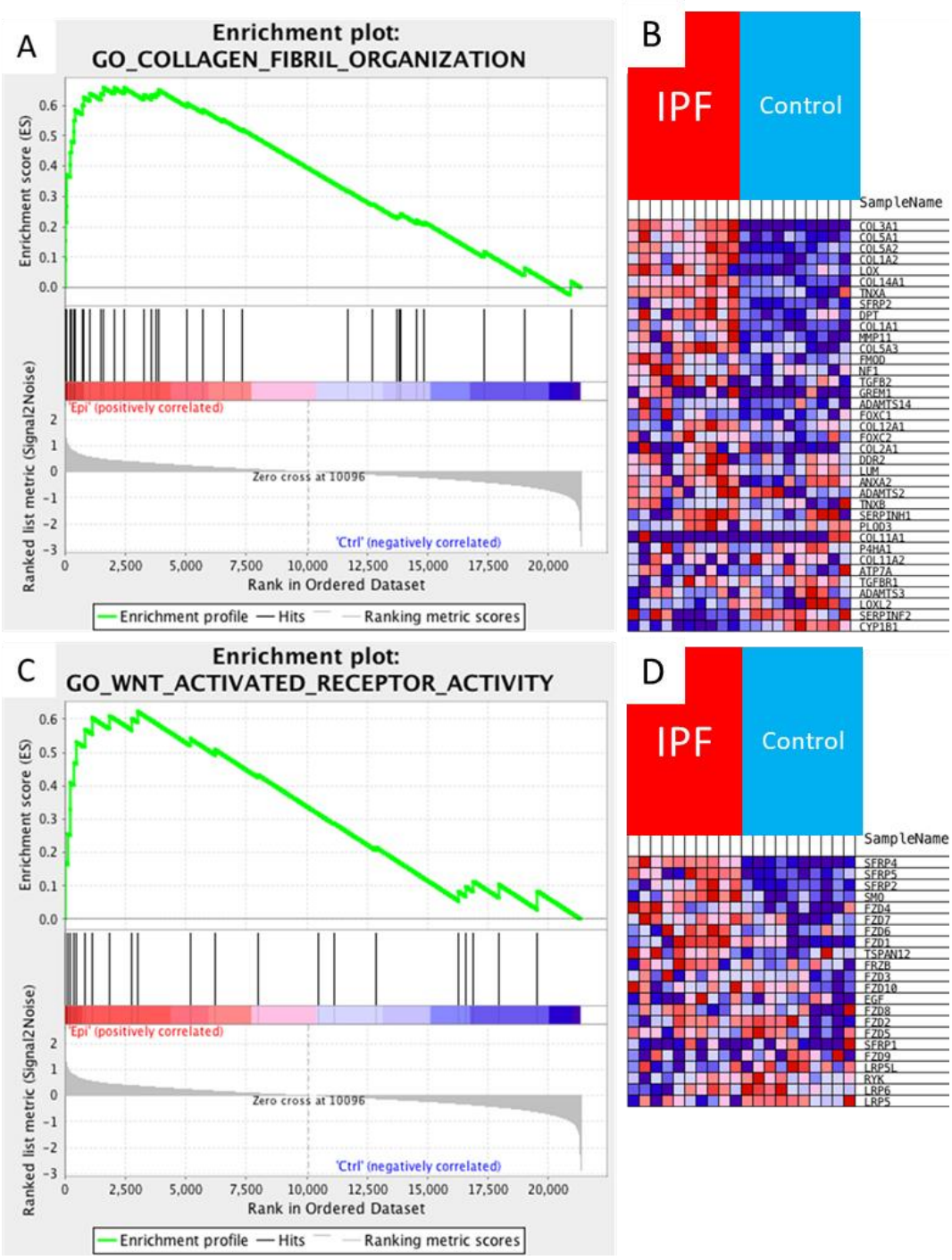


Figure 3.4: Plots showing enrichment of particular gene ontology terms in IPF alveolar septae LCMD data compared to control alveolar septae. A. and C. Enrichment plots showing enrichment of collagen fibril organisation and Wnt-activated receptor activity GO terms. B. and D. Heatmaps showing normalised FPKM values for gene expression of genes in the collagen fibril organisation and Wnt activated receptor activity GO terms. Enrichment plots show enrichment score calculated by iterating down the ranked gene list. Genes in the selected GO terms are shown as vertical lines representing their position in the ranked gene list.

3.10 WISP1 is principally expressed in the fibroblast foci

As WISP1 was not highly expressed in the IPF alveolar samples, the data were used to identify whether WISP1 is expressed in fibroblastic foci from IPF tissue. Therefore, WISP1 expression levels were compared across the three different samples, and WISP1 was shown to be most strongly expressed in the fibroblast foci (Figure 3.5). This contradicts previous research which suggested that WISP1 expression may localise to the alveolar epithelium in IPF(191), and suggests that further investigation of WISP1 function is likely to be best conducted using lung fibroblasts as a model(192).

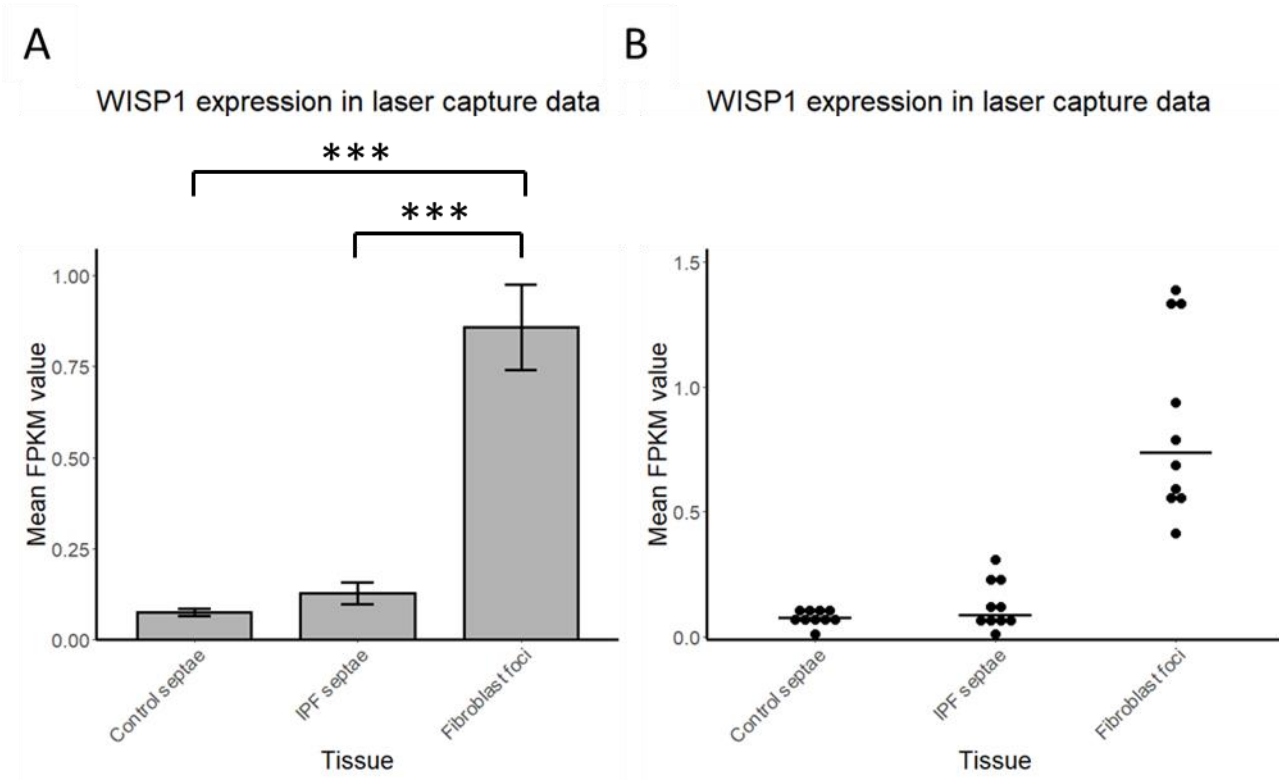


Figure 3.5: WISP1 expression in laser capture data. A. Bar chart showing mean expression of WISP1 in laser capture data. Error bars show standard error of the mean. WISP1 expression in fibroblast foci is significantly different to its expression in both IPF septae and control septae (t-test).

The currently accepted model of IPF pathogenesis suggests that the fibroblast focus is the tissue area that principally drives fibrosis, producing most of the collagenous extracellular matrix(193). Therefore, identifying enriched phenotypes in the fibroblast focus is important to identify the genes important for driving the disease.

3.11 GSEA shows enrichment for gene ontology terms associated with skeletal development in the fibroblast foci

In order to identify gene sets enriched in the fibroblast foci, GSEA comparing the fibroblast foci samples to the alveolar septae was performed. GSEA results are shown in Table 3.8. Table 3.9 shows enriched gene ontology terms associated with the fibroblastic foci LCMD RNAseq data. This identified a very strong enrichment for gene ontology terms associated with extracellular matrix production and skeletal development. The large numbers of gene ontology terms associated with skeletal development is consistent with the observation of bone-type collagen crosslinking in IPF lung tissue. An enrichment of genes associated with bone development in the fibroblast foci provides a potential mechanism for the increase in collagen stiffness seen in IPF. WISP-1 has also previously been associated with bone and cartilage development, suggesting that the production of bone and cartilage-like extracellular matrix present in IPF could be WISP-1 driven(150,194).

Table 3.8: top 25 enriched genes and gene descriptions ordered by GSEA ranking score in fibroblast foci compared to control alveolar septae. Gene set enrichment analysis was performed using the Broad institute's GSEA software. Signal2Noise metric was used for calculating gene rankings, with the default 1000 permutations.

Gene	Description	GSEA Ranking Score
COL15A1	Collagen type XV alpha 1 chain	4.17
COL14A1	Collagen type XIV alpha 1 chain	4.11
COL1A1	Collagen type I alpha 1 chain	3.94
COL1A2	Collagen type I alpha 2 chain	3.85
ANGPTL2	Angiopoietin like 2	3.56
CLMP	CXADR like membrane protein	3.55
MRVI1	Murine retrovirus integration site 1 homolog	3.51
FBN1	Fibrillin 1	3.45
HTRA1	Htra serine peptidase 1	3.41
ASPN	Asporin	3.33
PYCR1	Pyrroline-5-carboxylate reductase 1	3.33

THY1	Thy-1 cell surface antigen	3.15
OGN	Osteoglycin	3.10
FAM198B	Golgi associated kinase 1B	3.08
ROR2	Receptor tyrosine kinase like orphan receptor 2	3.08
GPX8	Glutathione peroxidase 8 (putative)	3.06
COL5A1	Collagen type V alpha 1 chain	3.06
HEPH	Hephaestin	3.04
COL3A1	Collagen type III alpha 1 chain	3.04
MMP2	Matrix metallopeptidase 2	2.99
SSC5D	Scavenger receptor cysteine rich family member with 5 domains	2.95
PCOLCE	Procollagen c-endopeptidase enhancer	2.94
TSHZ2	Teashirt zinc finger homeobox 2	2.93
GAL3ST4	Galactose-3-o-sulfotransferase 4	2.86
RCN3	Reticulocalbin 3	2.84
LRRC17	Leucine rich repeat containing 17	2.84
FBLN1	Fibulin 1	2.83
SERPINF1	Serpin family F member 1	2.83
IGFBP5	Insulin like growth factor binding protein 5	2.82
RASL11B	RAS like family 11 member B	2.81

Table 3.9: Top enriched gene ontology terms associated with fibroblast foci compared to control alveolar septae. Gene sets are ranked by normalised enrichment score, calculated using the GSEA software based on a ranked list of genes. Gene ranking was calculated using the Signal2Noise metric.

GO term	p-value	FDR q-value
Collagen binding	0.00	0.012046
Extracellular matrix component	0.00	0.01425
Dentate gyrus development	0.00	0.009785
Dermatan sulfate proteoglycan metabolic process	0.00	0.008445
Chondroitin sulfate proteoglycan biosynthetic process	0.00	0.008111
Ossification	0.00	0.007244
Extracellular matrix	0.00	0.006312
Cartilage development	0.00	0.006657
Proteinaceous extracellular matrix	0.00	0.006224
Connective tissue development	0.00	0.005936
Replacement ossification	0.00	0.005789
Chondrocyte differentiation	0.00	0.005612
Extracellular matrix structural constituent	0.00	0.005666
Extracellular structure organization	0.00	0.008809
Central nervous system projection neuron axonogenesis	0.00	0.008758
Wnt protein binding	0.00	0.009735
Sulfur compound catabolic process	0.00	0.009162
Endoplasmic reticulum lumen	0.00	0.008968
Embryonic cranial skeleton morphogenesis	0.00	0.008764
Proteoglycan metabolic process	0.00	0.009441
Sulfuric ester hydrolase activity	0.00	0.010086
Collagen fibril organization	0.00	0.010328
Multicellular organismal macromolecule metabolic process	0.00	0.009983
Proteoglycan biosynthetic process	0.00	0.010151
Keratan sulfate biosynthetic process	0.00	0.01012
Fibronectin binding	0.00	0.010588
Cellular response to amino acid stimulus	0.00	0.011157
Chondroitin sulfate biosynthetic process	0.00	0.010858
Basement membrane	0.00	0.011222
Embryonic skeletal system development	0.00	0.010878

Figure 3.6 shows the enrichment of genes associated with the ossification GO term. This includes several genes known to be enriched in IPF, including SPARC, as well as several matrix metalloproteinase genes, which have been shown to play a role in tissue remodelling in IPF(76,195).

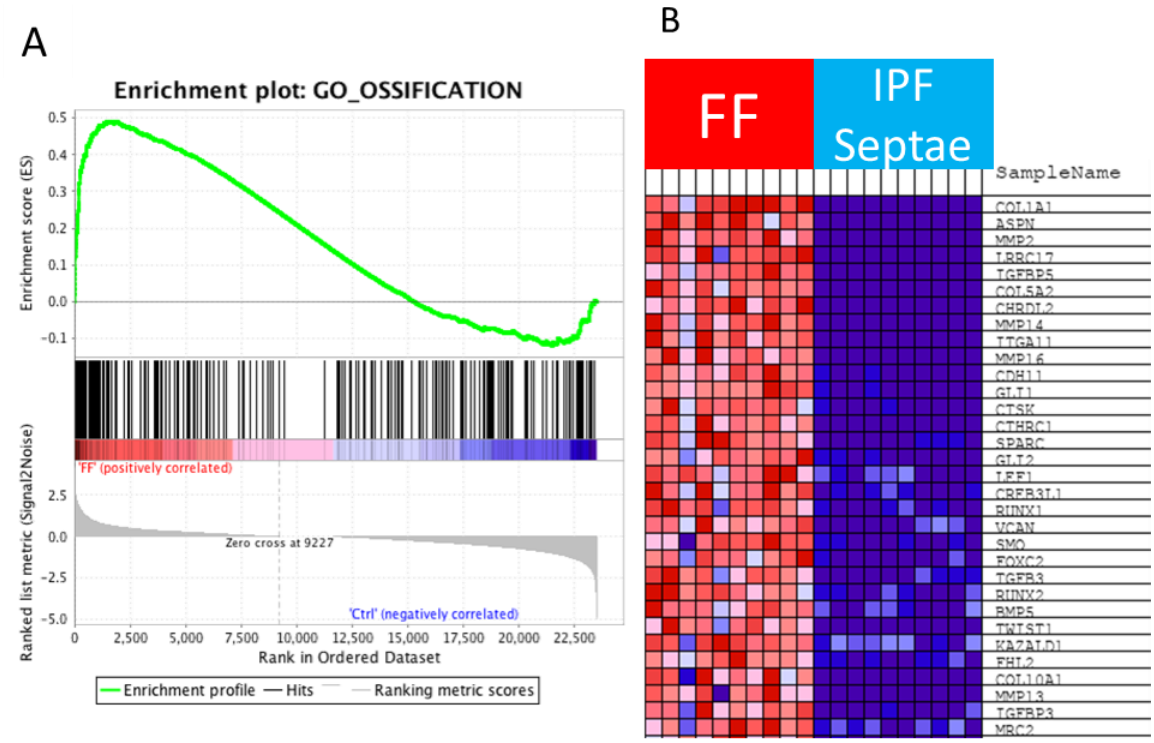


Figure 3.6: Enrichment of the Ossification gene ontology term in fibroblast foci (FF) compared to alveolar septae (IPF septae) samples. A. Enrichment plot showing the enrichment score for this GO term calculated by iterating down the ranked gene list. Genes in the selected GO term are shown as vertical lines representing their position in the ranked gene list. B. Heatmap showing normalised expression values for the most enriched ossification genes.

Figure 3.7 show genes associated with Wnt protein binding, another enriched GO term. Multiple Wnt receptor genes are upregulated, including genes coding for several frizzled proteins, the prototypical Wnt receptor(104). Interestingly, the most enriched gene in the fibroblastic foci samples in this GO term is ROR2, which codes for a receptor tyrosine kinase that has previously been implicated in bone development(196). Also upregulated is SFRP2, a Wnt antagonist that was

upregulated in IPF in the bulk transcriptomic datasets analysed above(197). Enrichment of proteins which bind Wnt ligands may suggest that cells in fibroblastic foci are more sensitive to Wnt signalling, potentially leading to increased WISP-1 production as a result.

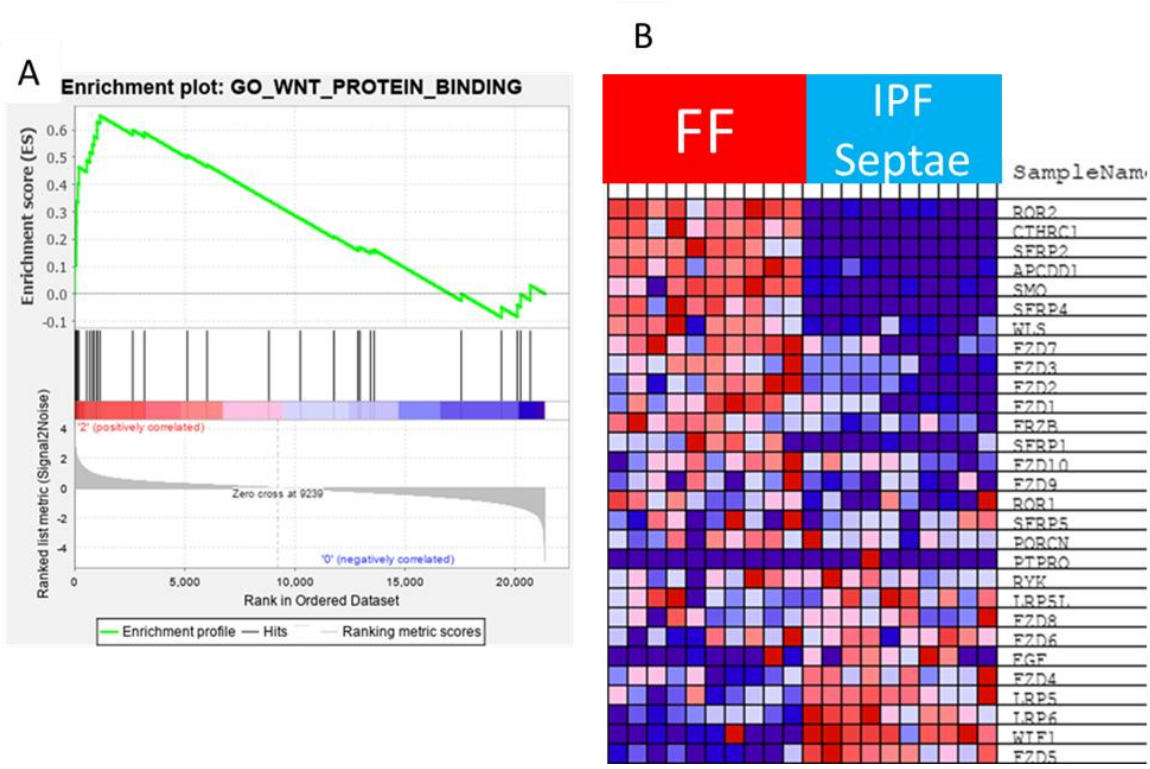


Figure 3.7: Enrichment of the Wnt protein binding gene ontology term in fibroblast foci (FF) compared to alveolar septae (IPF septae) samples. A. Enrichment plot showing the enrichment score for this GO term calculated by iterating down the ranked gene list. Genes in the selected GO term are shown as vertical lines representing their position in the ranked gene list. B. Heatmap showing normalised expression values for most enriched Wnt protein binding genes.

3.12 Single cell RNA-seq data allows a heterogeneous population of cells in healthy and diseased lungs to be defined.

Due to the nature of the above analysis comparing the fibroblast foci to the alveolar septae, it is difficult to parse any disease-specific effects due to differences in the two sample types. Fibroblast foci and alveolar septae have a different cellular makeup, making it difficult to identify

whether enriched gene sets and molecular phenotypes are due to disease-driven processes or differences between different tissue types.

IPF is a particularly problematic disease because of the extreme tissue heterogeneity seen in IPF, with areas of relatively normal looking lung epithelium close to fibrotic epithelium, and fibroblast foci. With this dataset, it was not possible to compare fibroblasts in IPF to fibroblasts in normal lung, as it was not possible to isolate comparable fibroblastic areas from normal lung, as fibroblastic foci are a pathologic feature of the disease.

This is further compounded by the complexity of fibroblast foci. Although often modelled as consisting solely of matrix-secreting myofibroblasts, they consist of fibroblasts and the overlying epithelium, and are likely a complex mixture of different cell types themselves. The same is even more true for alveolar septae – it consists of many different cell types, alveolar type I and type II cells, endothelial cells and immune cells.

In order to characterise different types of cells present in fibrotic and normal lungs more effectively, a previously published single-cell RNA-seq dataset was used. This dataset was generated using 30 samples from control and fibrotic lungs,10 control samples and 20 from patients with interstitial lung disease with the fibrosis samples including 12 patients with IPF, and 8 with other interstitial lung diseases, including sarcoidosis, nonspecific interstitial pneumonia (NSIP), chronic hypersensitivity pneumonitis and unclassifiable ILD(58).

Reasons for choosing this dataset include accessibility – the raw data is freely available online from the Gene Expression Omnibus (GEO) repository, unlike similar datasets, as well as many mesenchymal cells, particularly fibroblasts, identified in this dataset. As a primary purpose of this analysis is to identify functional roles of fibroblasts in IPF, a large sample of fibroblasts is important. At time of analysis, there was one other single-cell RNAseq dataset in pre-publication which analysed IPF tissue(198,199). This dataset was not selected for further analysis because the data was not accessible until final publication in mid-2020 (data was made available on GEA via accession number GSE136831 in May 2020. It also identified a smaller proportion of cells as mesenchymal, so is less suitable for the characterisation of different mesenchymal cell and fibroblast populations.

The study used 10x Chromium sequencing, 114,396 cells were sequenced and identified as having reads of high enough quality for further analysis.

Figure 3.8 A shows a Uniform Manifold Approximation and Projection (UMAP) dimensional reduction plot for this dataset, with different cell types colour-coded and labelled. This

dimensional reduction technique allows visualisation of different clusters of cells, with similar cells closer together on the plot, and distances between clusters being a useful heuristic for the variation between different clusters in the group. This UMAP plot was generated from the data normalised using the Seurat R package, and coordinates were generated using the University of Southampton High Performance Computing centre's IRIDIS 4 compute cluster. Clustering and cell type identification was performed by the original paper authors. Notably, their clusters map extremely closely to discrete cell clusters on the UMAP plot, suggesting both that UMAP is an appropriate way to capture dataset variation, and supporting their cell type assignments in the original publication.

While this analysis was previously performed in the original paper, it was redone here firstly so that new visualisations of the data could be generated, for example to produce graphs displaying the expression levels of genes of interest in this data for example. In order to generate UMAP and t-SNE visualisations, data had to be re-analysed to determine the coordinates associated with each visualisation. While the exact parameters for this analysis were not available at the time of this analysis, the broad methodology, using Seurat to identify highly variable genes, then performing principal components analysis as an initial dimensional reduction step before performing further dimensional reduction, was. Therefore, this analysis approach was followed. The UMAP plots generated were then mapped back onto the clusters identified in the paper. Unlike the dimensional reduction co-ordinates, these clusters were included in the metadata available from GSE135893. The high colocalisation of this mapping to the UMAP plots shown in the original paper also provides a useful validation of the dimensional reduction used.

Figure 3.8 B shows the same UMAP plot with final diagnosis as the colour coding. While most cell types are present in both control and ILD samples, there are some clear differences between the two states. Control and ILD cells tend to group separately within the larger cell type clusters – particularly evident in alveolar type II cells, endothelial cells and macrophages, while certain cell types appear to be disease specific – KRT5-/KRT17+ and PLIN2 + and HAS1 high fibroblasts appear largely IPF-specific, for example.

There is clearly a great deal of heterogeneity in the cell types present in this dataset, reflecting the diversity of cells present in the lung. In contrast to the laser-capture microdissection dataset, this single-cell data captures that heterogeneity better. However, it gives no localisation data as to where these cell types are found. Combining the granular, single-cell level sequencing data with the highly localised laser-capture microdissection sequencing data would give a comprehensive picture of the cell types present in different types of IPF tissue. Characterising

these cell types from the single cell RNAseq dataset would then give an insight into how these cell types behave in IPF.

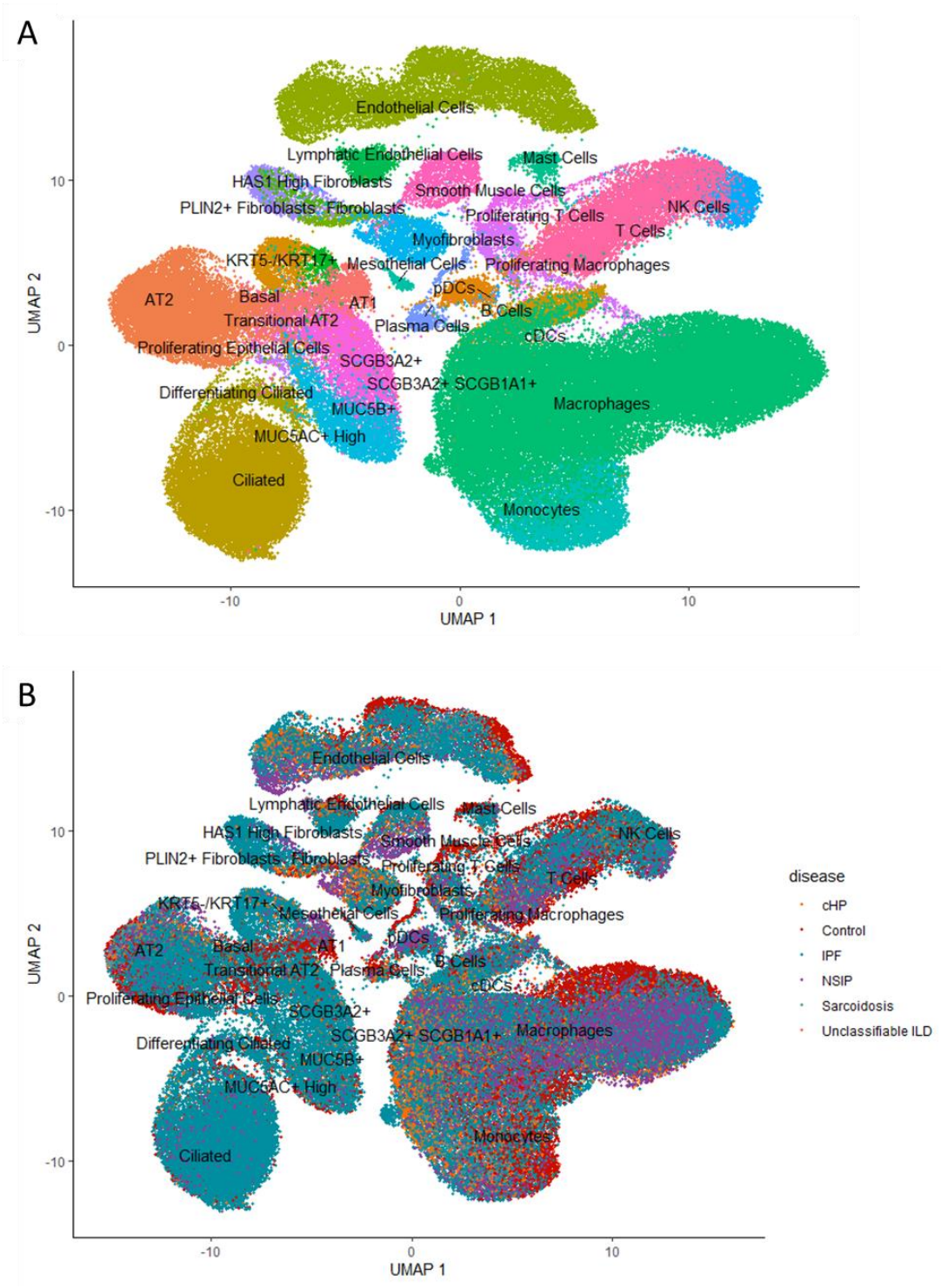


Figure 3.8: Distribution of different cell types and interstitial lung diseases for Single cell RNA-seq data in Habermann et al. Data are presented as a Uniform Manifold Approximation and projection (UMAP) plot, generated using the Seurat R package. A. UMAP plot showing different cell types in this dataset. Cell identifications were taken from the single cell RNA-seq metadata. B. UMAP plot showing different ILDs and control cells present in this dataset.

3.13 CibersortX, an *in silico* cell sorting algorithm, allows genetic signatures associated with different cell types to be determined

In order to deconvolute the laser capture microdissection RNA-seq data, the *in silico* cell sorting algorithm CibersortX was used. This takes as its input an n x m matrix where n is genes and m is individual cells. These cells are grouped according to a user-defined phenotype, in this case the predefined cell type identified in the original paper.

90 cells were taken from each of the 31 cell types defined in the Kropski single cell dataset for a total of 2760 cells from the whole dataset. Cells were only selected from IPF and control samples, with the other ILDs in the dataset excluded. Cells were then randomly selected from within each of their associated cell type lists. From this input data, a cell signature matrix was determined, identifying specific gene expression patterns associated with individual cell types. Between 300 and 500 genes were identified as a molecular barcode defining each cell type. Figure 3.9 shows a heatmap of the different barcodes associated with each individual cell type, showing a distinct, cell specific gene signature is identifiable for all cell types.

Having determined that it is possible to identify a cell-type specific gene signature from these data, and that these signatures reflect previously known genetic patterns associated with the appropriate cell type, they were applied to the laser-capture RNA-seq data to identify the proportion of different cell types present in the different laser-capture RNA-seq samples.

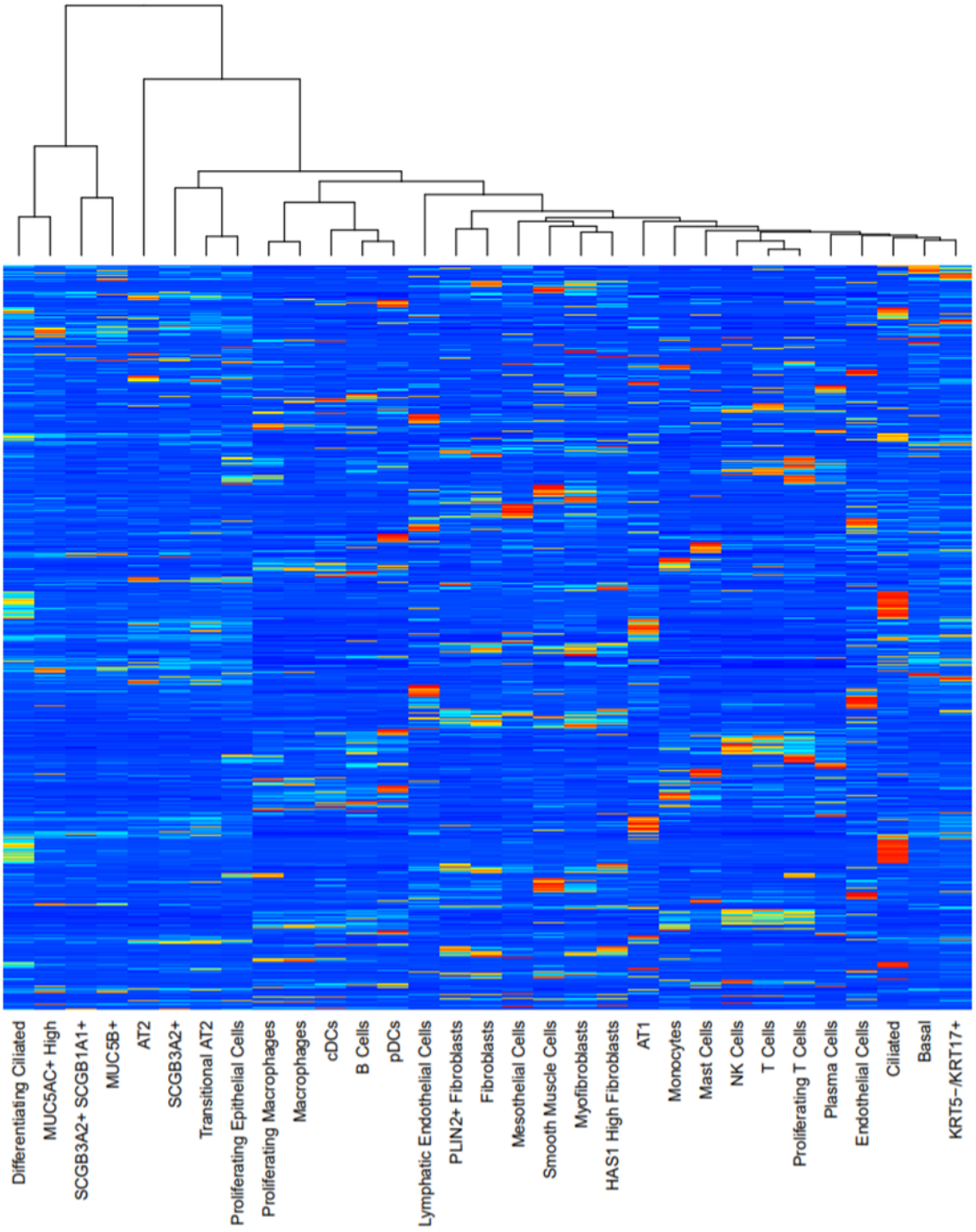


Figure 3.9: Heatmap showing gene signatures associated with different cell types. Generated using CibersortX.

3.14 CibersortX analysis of LCMD RNAseq data reveals different cellular compositions for different sample types

Parameters for the CibersortX analysis were as follows: batch correction was applied as the laser-capture microdissection samples were input as fragments per kilobase per million mapped reads (FPKM) values, while the single cell RNA-seq signature matrix was calculated using raw counts as the input. All other parameters were kept to the default values.

Figure 3.10 shows the combined proportions calculated using this CibersortX analysis, with cell types grouped as epithelial, endothelial, mesenchymal or immune according to their classification in table. Broadly, the fibroblast foci samples show enrichment in mesenchymal cell types, as might be expected given the known functional role of fibroblast foci in IPF, while the two alveolar septae samples have similar proportions of the different cell types.

Figure 3.11 shows the breakdown by sample of this analysis, while Figure 3.12 shows the mean proportion of each cell type by sample. Notably, it is clear from Figure 3.11 that two of the IPF alveolar septae samples have significantly different cellular compositions than others. These two samples correspond to the samples on the PCA plot in figure which cluster away from the rest of the IPF alveolar septae samples. As the laser-capture microdissection data combined multiple alveolar septae samples from the lung tissue of each patient, it is likely that this is not an artefact due to sampling of different parts of the lung tissue, but a different manifestation of the disease in these patients. There is a larger proportion of airway type cells, possibly indicating a later stage of disease where honeycomb cysting has removed the majority of the alveolar epithelial structure in the lungs.

Identifying trends in the cell type specific data, rather than the broader categories of epithelial, endothelial, immune and mesenchymal between different cell types allows a more granular look at the data, made possible by using single cell RNA-seq data and the large number of different cell types identified in this dataset. In particular, there is a strong enrichment in myofibroblasts and HAS1-high fibroblasts in the fibroblast foci compared to the two alveolar septae samples.

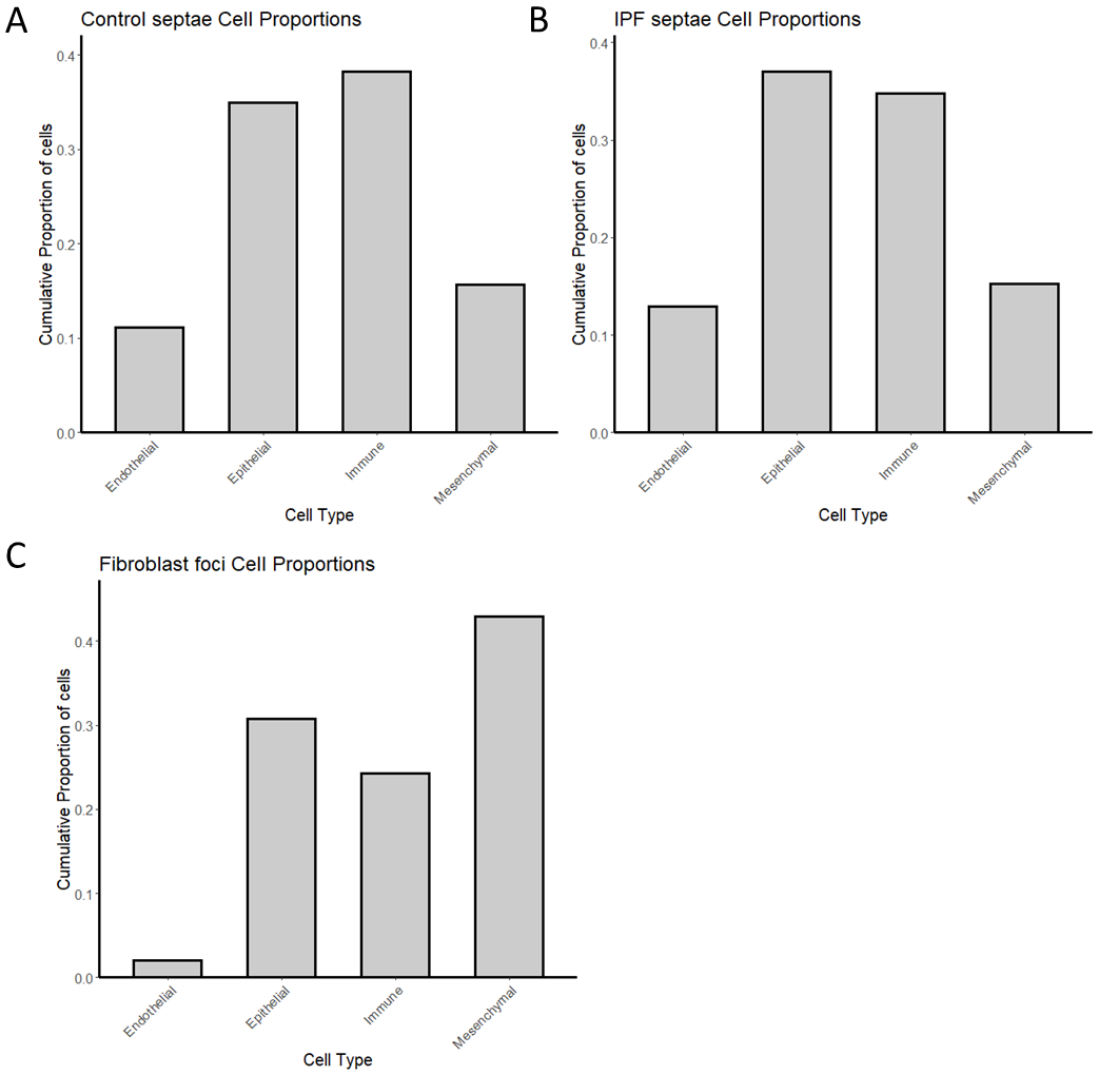


Figure 3.10: Charts showing cumulative values of the means of different cell types grouped by cell classifier in three different lung tissue types. Mean proportion of the cell types shown in Figure 3.12 classified as endothelial, epithelial, immune and mesenchymal were summed to give a cumulative mean proportion for each cell classifier. A. Control septae. B. IPF septae. C. Fibroblast foci.

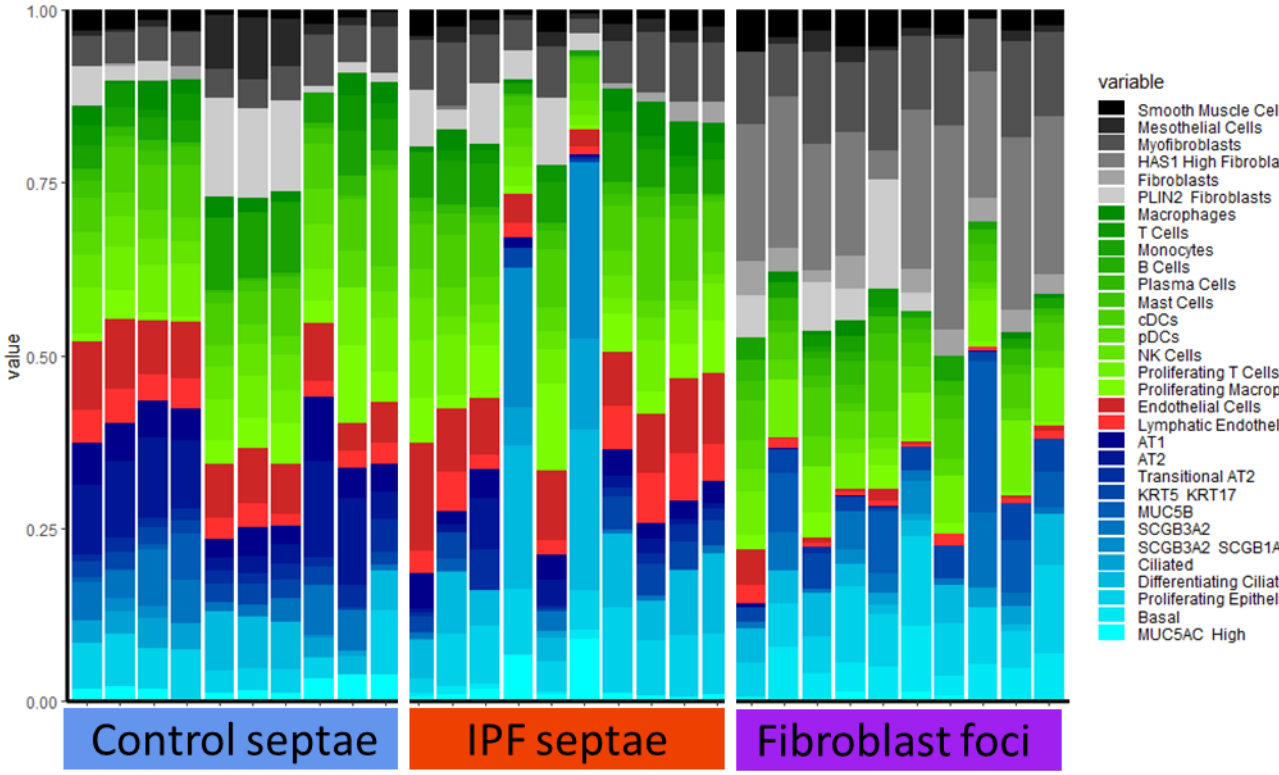


Figure 3.11: Relative proportions of different cell types in laser-capture data from CibersortX analysis. 90 randomly chosen cells from control and IPF single cells were selected and used to generate cell type-specific gene signatures. These were then used to estimate prevalence of cell types for individual samples in the laser capture microdissection RNA-seq dataset.

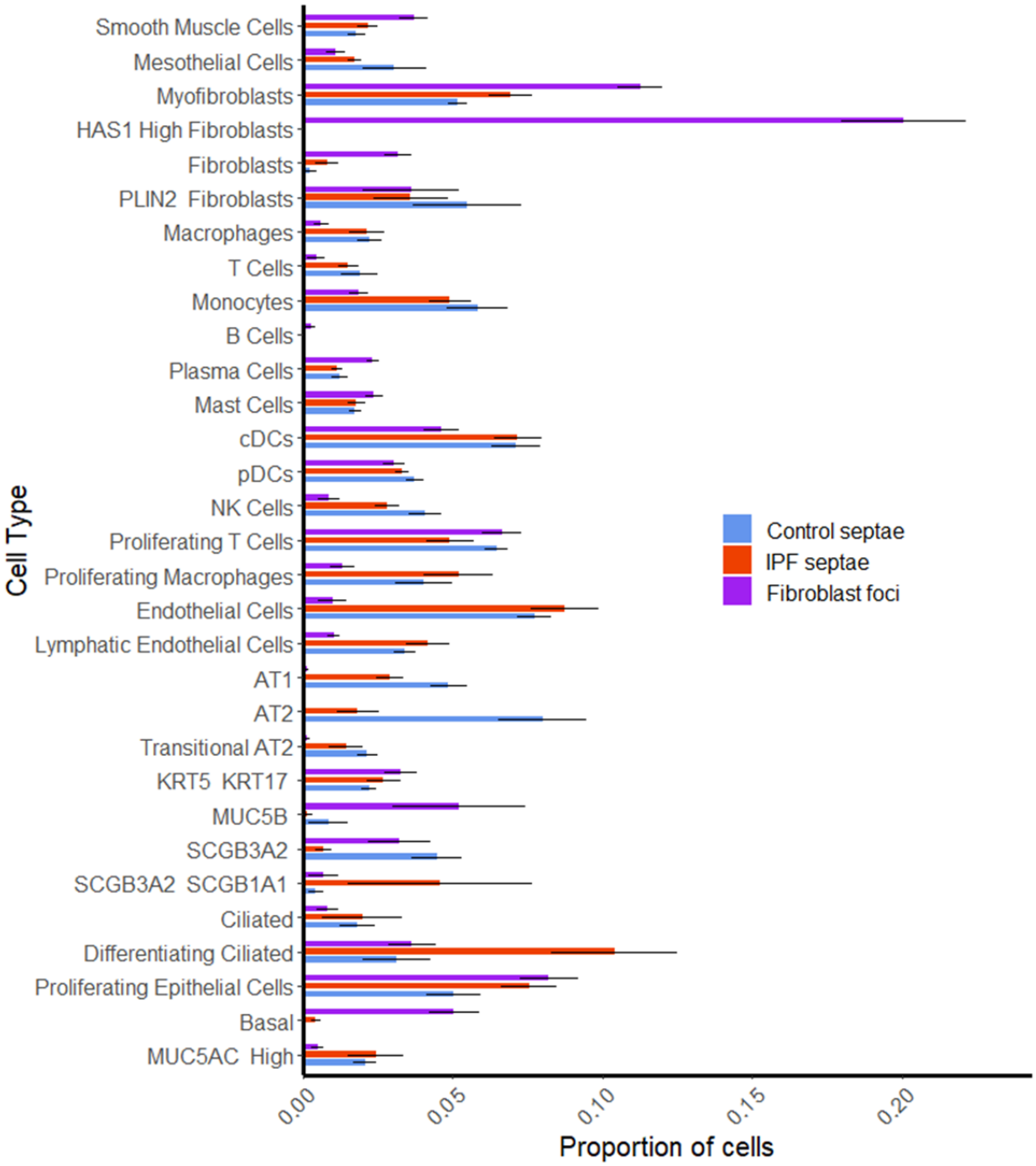


Figure 3.12: Chart showing mean proportion of cells in different tissue types for all identified cell types in the Banovich/Kropski dataset according to CibersortX analysis. Error bars are standard error of the mean.

3.15 TSNE plots of different mesenchymal cells show segregation of fibroblast subtypes, as well as IPF-specific fibroblast subtypes.

To characterise the different mesenchymal cells in this dataset, dimensional reduction was performed on the single cell RNA-seq data to show the distribution of the different mesenchymal cell types, and, importantly, how well they segregate on a dimensional reduction plot based on their assigned cluster identities. Figure 3.13 shows T-stochastic nearest neighbour embedding (TSNE) dimensional reduction plots showing the distribution of mesenchymal cells in the Banovich/Kropski dataset. Figure 3.13 A shows the distribution of cell types for this data, while Figure 3.13 B shows the disease distribution. The cell types segregate well, clustering together, suggesting that the majority of the variation in this dataset is due to the different types of cells present in the lung tissue used for this dataset. However, as the disease distribution shows for myofibroblasts and fibroblasts, there is less segregation between IPF, other ILDs and control lungs. Although the classical picture of fibroblasts in IPF is of aggregates of highly active, secretory myofibroblasts, there is little evidence of a qualitative difference between myofibroblasts in IPF compared to normal lung, although there may be a quantitative difference(50).

The disease plot does show that most PLIN2-positive and HAS1-high fibroblasts are from patients with IPF, with a small cluster of them associated with non-specific interstitial pneumonia (NSIP). This is of interest, as these fibroblasts are strongly enriched in the fibroblast foci in the CibersortX analysis, with HAS1-high fibroblasts being particularly enriched. The role these cells play in IPF is therefore very interesting and may provide new insight into the functional roles of different fibroblast subtypes in IPF.

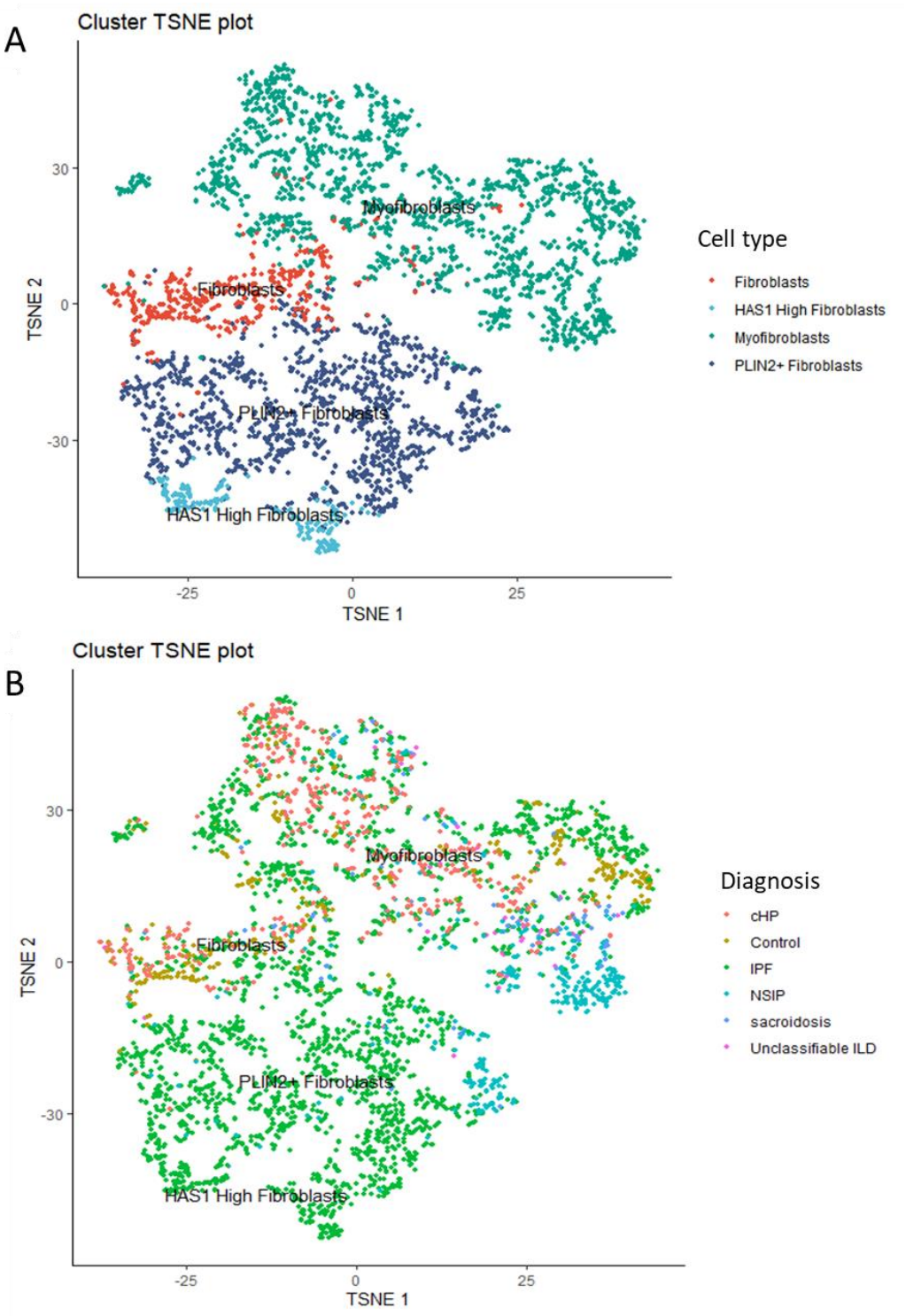


Figure 3.13: T-distributed stochastic nearest neighbour embedding (TSNE) dimensional reduction plots showing distribution of fibroblasts in the Banovich/Kropski single cell RNA-seq dataset. Fibroblast types and diagnoses were taken from the original paper, TSNE values were calculated using the Seurat R package. A. Distribution of fibroblast subtypes. B. Distribution of diagnoses.

3.16 Characterising enriched genes in myofibroblasts compared to other fibroblast subtypes shows a strong enrichment for pro-fibrotic genes

Having identified the proportions of cells in different parts of the dataset, in order to characterise these cell types, it is important to identify gene expression signatures associated with them. In order to do this, the Seurat FindMarkers() function was used. This uses by default a Wilcoxon Rank Sum test to compare two different groups of samples. In order to identify distinguishing gene signatures from individual fibroblast subtypes, they were compared to all the other mesenchymal cell types in the dataset. Including non-mesenchymal cells in the analysis would be problematic, as common gene expression signatures within all fibroblast subtypes would be difficult to parse out from the cell type-specific signatures when comparing with to non-mesenchymal cell types.

Initially, myofibroblast associated markers were identified. Myofibroblasts are a cell type with a very well-defined gene expression signature associated with them, including α smooth muscle actin and vimentin gene expression. Table 3.10 shows the marker genes associated with the myofibroblasts in this single-cell RNA-seq dataset. Genes strongly associated with myofibroblasts in the literature such as ACTA2 and VIM are strongly associated with this phenotype, as well as several known pro-fibrotic genes such as ACTA2 (α-SMA) and POSTN (periostin).

In order to identify the molecular phenotype of these cells, the online gene ontology enrichment tool GOrilla was used. This gives a hierarchical tree of gene ontology terms associated with cellular processes, functions and components, where more broadly applicable gene ontology terms containing many genes are at the top of the tree, with progressively more specific terms with a smaller number of genes at each subsequent level.

The treemap shown in Figure 3.14 shows the enriched GO terms associated with myofibroblasts. Terms associated with extracellular matrix development are highly enriched, strengthening the known role of these cells as highly secretory ECM producers.

Table 3.10: Top genes associated with myofibroblasts in this dataset. FindMarkers() function in Seurat used with Wilcoxon rank sum test to identify differentially expressed genes in myofibroblasts compared to all other fibroblast types. P values are adjusted with the Bonferroni multiple test correction using all genes in the dataset.

Gene Name	p val	avg logFC	pct.1	pct.2	p val adj
A2M	8.46E-149	1.96	0.92	0.61	2.85E-144
POSTN	6.65E-85	1.78	0.41	0.08	2.24E-80
LTBP2	6.70E-211	1.64	0.74	0.15	2.26E-206
ASPN	1.28E-82	1.58	0.47	0.14	4.30E-78
ACTA2	3.50E-45	1.58	0.53	0.34	1.18E-40
FN1	1.52E-156	1.56	0.97	0.92	5.12E-152
TAGLN	1.77E-86	1.54	0.67	0.32	5.96E-82
DKK3	1.01E-203	1.48	0.73	0.15	3.41E-199
TNC	3.43E-125	1.44	0.41	0.03	1.16E-120
LIMCH1	2.55E-144	1.43	0.61	0.15	8.58E-140
COMP	1.47E-70	1.36	0.29	0.04	4.97E-66
BGN	5.07E-138	1.36	0.89	0.65	1.71E-133
COL8A1	4.65E-101	1.34	0.66	0.29	1.57E-96
PPP1R14A	1.39E-150	1.33	0.53	0.07	4.68E-146
ITGBL1	5.11E-98	1.30	0.41	0.07	1.72E-93
LBH	5.69E-124	1.29	0.67	0.28	1.92E-119
CD82	2.24E-179	1.26	0.57	0.06	7.56E-175
C10orf10	2.48E-63	1.24	0.41	0.12	8.35E-59
CTGF	4.70E-42	1.23	0.60	0.39	1.59E-37
CES1	3.35E-119	1.23	0.65	0.21	1.13E-114

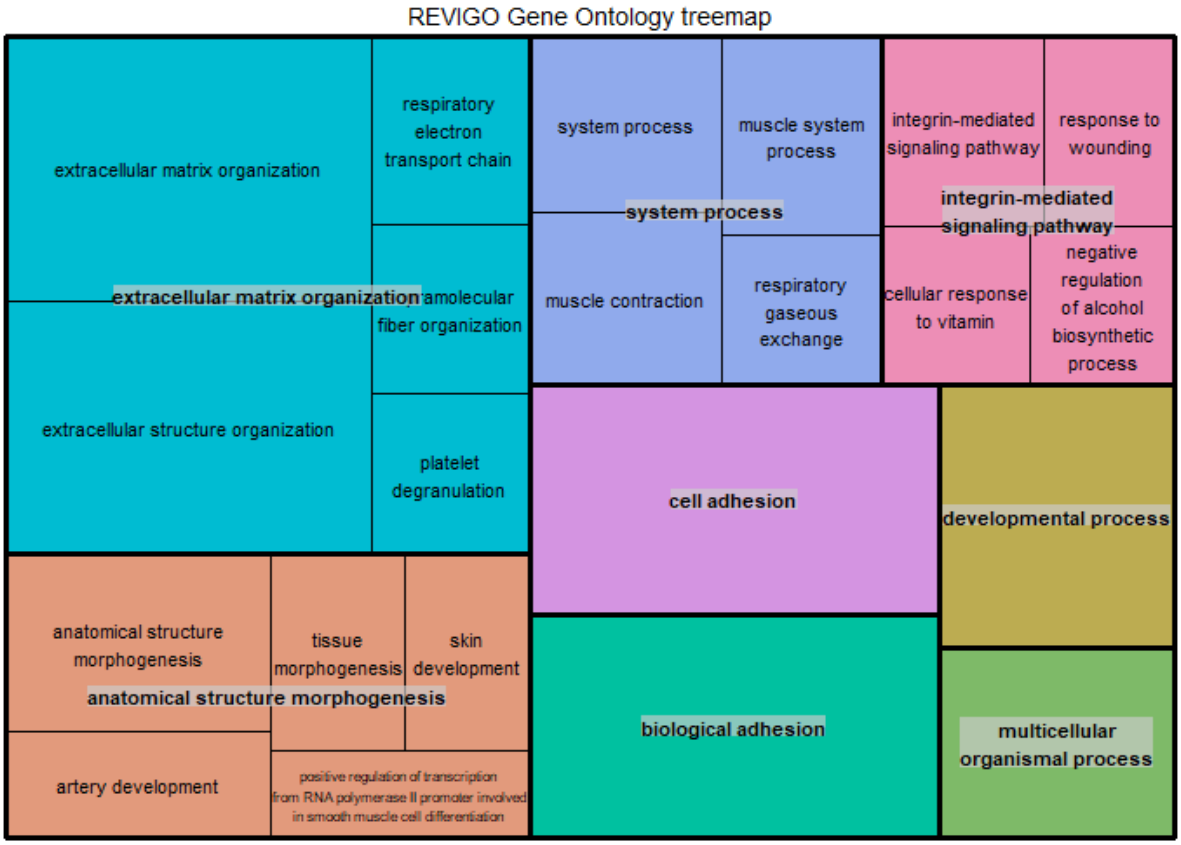


Figure 3.14: Treemap showing the top gene ontology terms associated with enriched myofibroblast genes. GO enrichment was performed in GOrilla, with enriched ontology terms calculated using a target list of marker genes compared to a background list of all genes in the dataset. Size of squares is proportional to the significance of the GO enrichment, and inset squares of the same colour are associated with the same overarching GO term.

3.17 HAS1-high fibroblasts have a characteristic stress response and show markers associated with senescence

Table 3.11 shows the top 20 upregulated genes associated with the HAS1 fibroblasts, ordered by log fold change. The most upregulated gene is SERPINB2, or plasminogen activator 2 (PAI-2). This is a serine protease inhibitor which inhibits urokinase and may be associated with sensing redox potential(200). In fibroblasts, it has also been associated with senescence(201). Overactive telomerases are associated with an increased likelihood of developing IPF, and epithelial senescence has previously been identified as being a hallmark of the disease. Several other genes, particularly the chemokines CXCL8, CXCL2 and CXCL3 are associated with the senescence

associated secretory phenotype, where senescent cells secrete pro-inflammatory cytokines, growth factors and proteases(202,203). A subunit of one of the drivers of this SASP, NF-κB, NFKBA1, is also significantly upregulated, albeit with a logFC of 0.32, lower than the other factors identified above(204).

Figure 3.15 shows the enriched gene ontology terms associated with this set of genes. Although senescence-related gene ontology terms are not identified, there are many GO terms associated with response to stimuli, including heat and chemical stress, and oxygen-containing compounds. These cells exhibit a very active stress response, which may be a result of their senescence. Other enriched GO terms include many terms associated with cell cycle regulation, another suggestion that these cells are no longer dividing.

HAS1, or hyaluronan synthase 1, was used as a defining marker for these cells in the original paper. HAS1 is an enzyme which produces hyaluronic acid, an extracellular polysaccharide associated with cartilage, which is often upregulated in wound-healing(205). As current hypotheses of IPF pathogenesis associate it with an abnormal, chronic wound-healing response, upregulation of HAS1 makes sense within the context of the disease and serves to distinguish these cells as stress-responding drivers of that wound-healing response(193).

According to the CibersortX analysis above, the dominant cell types in fibroblast foci are myofibroblasts and HAS1 high fibroblasts, suggesting that these HAS1 high fibroblasts may play an important role in the fibroblast foci. They are not the predominant expressors of extracellular matrix components, with myofibroblasts producing more collagen according to this dataset – although they do still produce ECM components (see below). Instead, as well as producing ECM components directly, their highly secretory, senescence associated phenotype may be associated with prolonging the chronic wound-healing response characteristic of IPF pathogenesis.

Table 3.11: Top genes associated with HAS1-high fibroblasts in this dataset. FindMarkers() function in Seurat used with Wilcoxon rank sum test to identify differentially expressed genes in HAS1-high fibroblasts compared to all other fibroblast types. P values are adjusted with the Bonferroni multiple test correction using all genes in the dataset.

Gene Name	p val	avg logFC	pct.1	pct.2	p val adj
SERPINB2	5.66E-52	1.80	0.26	0.02	1.91E-47
CXCL8	5.93E-52	1.72	0.51	0.11	2.00E-47
KIAA1217	8.25E-146	1.55	0.96	0.18	2.78E-141
LINC00152	1.68E-72	1.55	0.93	0.39	5.67E-68
HMOX1	8.06E-88	1.52	0.84	0.21	2.72E-83
PTGS2	2.59E-74	1.52	0.76	0.20	8.71E-70
TM4SF1	1.69E-46	1.46	0.61	0.18	5.70E-42
SGK1	2.97E-79	1.42	0.96	0.36	1.00E-74
ARC	2.42E-61	1.34	0.74	0.21	8.15E-57
CXCL2	1.44E-47	1.29	0.83	0.30	4.85E-43
DKK1	5.11E-62	1.28	0.52	0.09	1.72E-57
HSPD1	1.01E-49	1.27	0.99	0.69	3.39E-45
IER3	2.99E-50	1.27	0.96	0.52	1.01E-45
CXCL3	2.45E-31	1.26	0.47	0.14	8.26E-27
HSPH1	7.23E-61	1.24	0.99	0.53	2.44E-56
MIR4435-2HG	3.31E-63	1.21	0.88	0.35	1.12E-58
UGDH	2.22E-65	1.20	0.99	0.59	7.48E-61
PDLIM3	2.18E-47	1.20	0.96	0.57	7.35E-43
HAS1	2.24E-65	1.19	0.97	0.43	7.54E-61

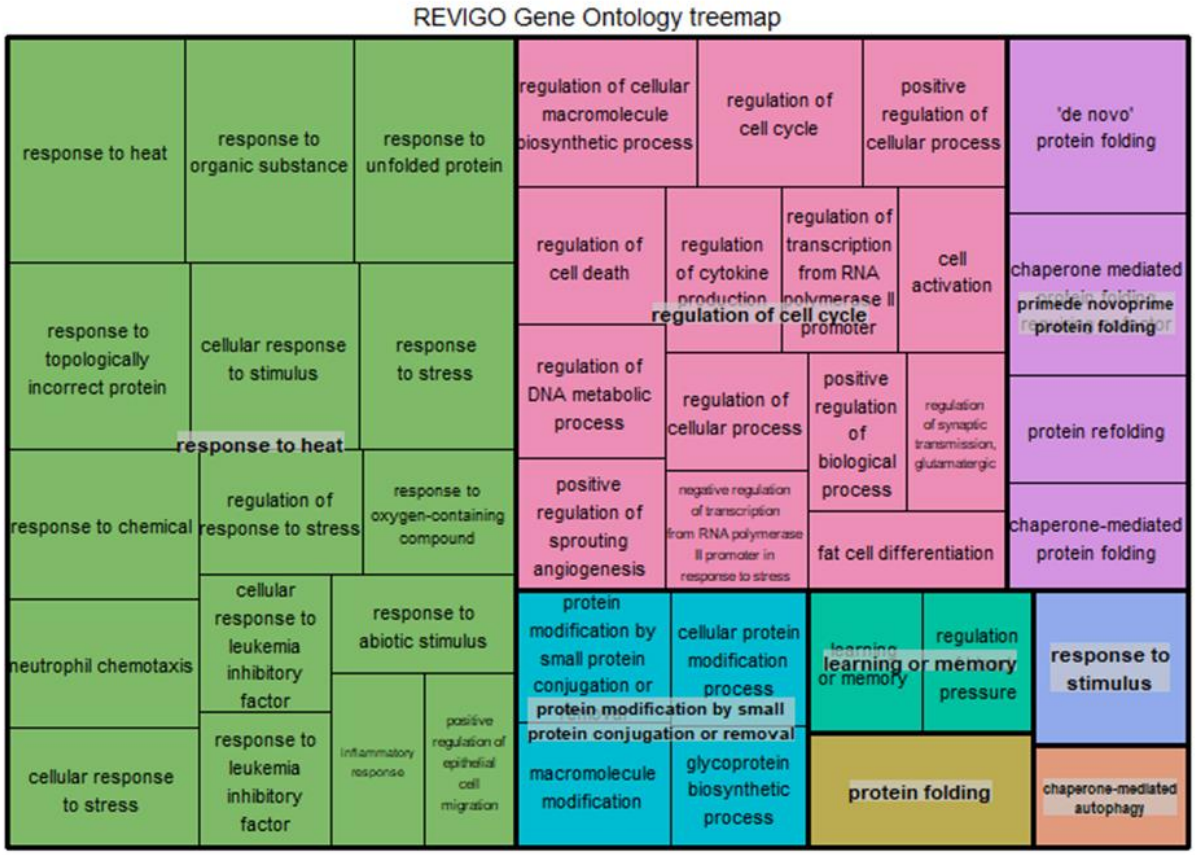


Figure 3.15: Treemap showing the top gene ontology terms associated with enriched HAS1-high fibroblast genes. GO enrichment was performed in GOrilla, with enriched ontology terms calculated using a target list of marker genes compared to a background list of all genes in the dataset. Size of squares is proportional to the significance of the GO enrichment, and inset squares of the same colour are associated with the same overarching GO term.

3.18 PLIN2 positive fibroblasts have a gene signature associated with metal ion sensing and metallothionein gene expression

The final fibroblast subtype, PLIN2 positive fibroblasts, make up a small proportion of cells in the CibersortX analysis. However, they represent a large proportion of the cells isolated from the IPF patients in this study. Figure shows they are also almost entirely IPF specific, not being found in the other ILDs in this study or the control cells. Their role in IPF is therefore of interest, as a fibroblast population which is disease-specific, but which is underrepresented in the fibroblast

foci, previously identified as the major fibroblast-containing population in this disease. However, it is worth noting that a PLIN2-positive fibroblast signature was identified in the control samples in the CibersortX analysis above, suggesting that this cell type may not be IPF specific.

PLIN2 is perilipin-2, a protein important for the formation of lipid droplets in adipocyte and other tissues. In mice, it is a known marker for lipofibroblasts(206). These cells are well-attested in mice, but controversial in humans(57), and are associated with maintenance of the epithelial barrier, and renewal. Lipofibroblasts are peroxisome proliferator-activated receptor gamma (PPARγ) expressing cells which provide lipids to ATII cells, allowing them to produce surfactant, and provide a protective effect against oxidant-induced epithelial damage(54). The possibility that PLIN2 positive fibroblasts in IPF have a role analogous to lipofibroblasts in mice is intriguing – they would represent a possible mechanism by which IPF sufferers could mitigate the epithelial disruption seen in disease.

As with the other fibroblast types described above, Table 3.12 shows the upregulated genes associated with PLIN2 positive fibroblasts. A clear signature associated with metallothionein genes (prefixed MT-, e.g. MT1A) is present. Metallothioneins are cysteine rich proteins which bind divalent metal ions, such as Zn^{2+} using the thiol groups on their cysteine residues(207). Functionally, metallothioneins are important in protection of cells and tissues against oxidative stress due to the redox sensitivity of their thiol groups, and metal toxicity, being able to sequester both physiological and non-physiological metal ions(208). The enrichment of metallothionein gene expression, as well as other genes which are sensitive to oxidation state and defend against cellular stress, such as HMOX1(209), suggests that PLIN2 positive fibroblasts may play a role in mitigating the ECM producing, fibrotic phenotypes of other fibroblast types in IPF.

Table 3.12: Top genes associated with PLIN2-positive fibroblasts in this dataset. FindMarkers() function in Seurat used with Wilcoxon rank sum test to identify differentially expressed genes in PLIN2-postive fibroblasts compared to all other fibroblast types. P values are adjusted with the Bonferroni multiple test correction using all genes in the dataset.

Gene Name	p val	avg logFC	pct.1	pct.2	p val adj
MT1A	4.22E-171	2.08	0.76	0.23	1.42E-166
PLA2G2A	2.34E-192	1.97	0.83	0.23	7.88E-188
FST	9.83E-110	1.80	0.61	0.20	3.31E-105
PTX3	1.22E-68	1.55	0.42	0.11	4.12E-64
MT1M	1.42E-92	1.53	0.73	0.38	4.77E-88
MT2A	7.66E-167	1.33	0.99	0.90	2.58E-162
PLIN2	2.40E-105	1.29	0.67	0.26	8.09E-101
ERRFI1	2.48E-184	1.28	0.90	0.34	8.36E-180
C3	7.28E-234	1.28	1.00	0.65	2.45E-229
CTSL	4.05E-164	1.23	0.96	0.61	1.36E-159
MYC	6.62E-187	1.23	0.92	0.33	2.23E-182
UAP1	4.34E-203	1.21	0.93	0.39	1.46E-198
ACSL4	1.44E-137	1.19	0.79	0.31	4.86E-133
PLAU	4.22E-82	1.16	0.63	0.27	1.42E-77
BDKRB1	4.36E-97	1.13	0.49	0.11	1.47E-92
GFPT2	1.00E-190	1.11	0.90	0.31	3.38E-186
TFPI2	9.24E-77	1.10	0.46	0.12	3.11E-72
IGF2	3.80E-177	1.09	0.82	0.21	1.28E-172
NAMPT	1.32E-157	1.08	0.89	0.42	4.44E-153

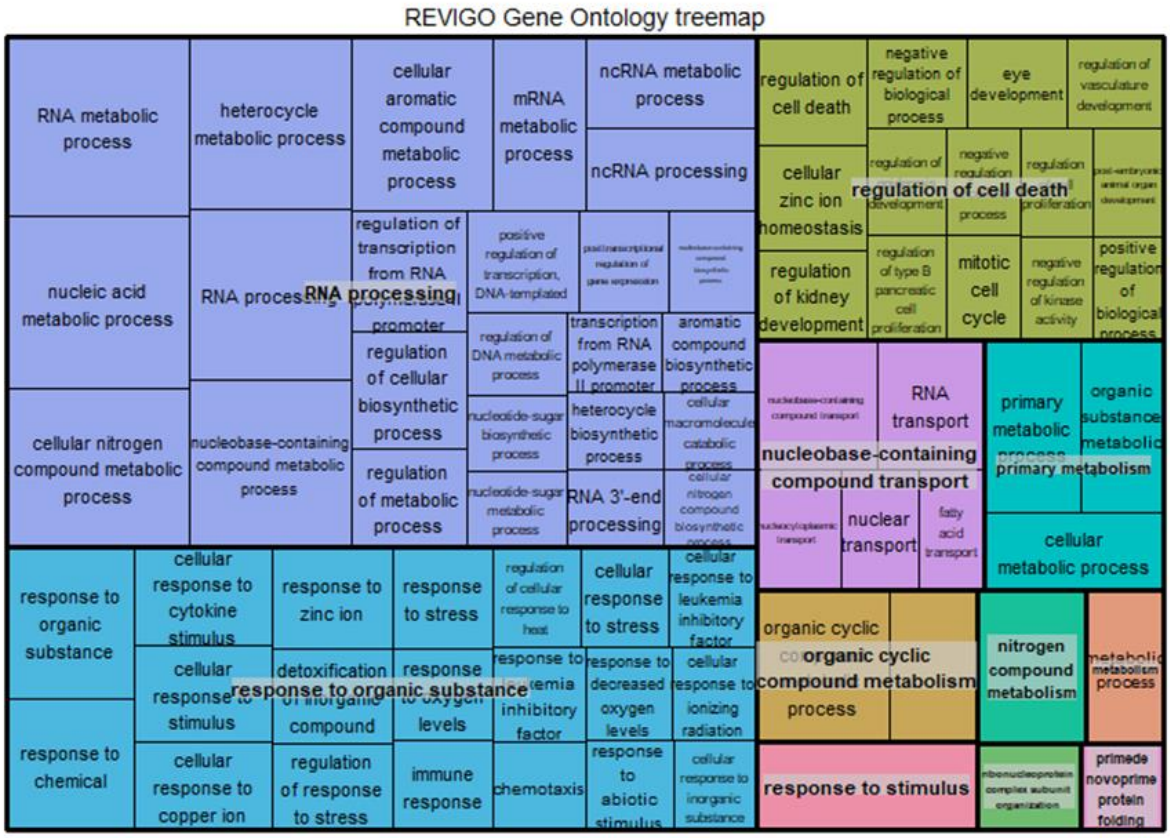


Figure 3.16: Treemap showing the top gene ontology terms associated with enriched PLIN2-positive fibroblast genes. GO enrichment was performed in GOrilla, with enriched ontology terms calculated using a target list of marker genes compared to a background list of all genes in the dataset. Size of squares is proportional to the significance of the GO enrichment, and inset squares of the same colour are associated with the same overarching GO term.

3.19 Despite shared marker gene expression, mouse lipofibroblasts and PLIN2 positive fibroblasts do not have much in common

In order to identify similarities between mouse lipofibroblasts and PLIN2 positive fibroblasts, a mouse fibroblast single-cell RNA-seq dataset which contains a clear, well defined population of lipofibroblasts was used. This dataset identified 1,943 fibroblasts from normal mouse lung and 3,386 from fibrotic mouse lung, these mice having been treated with the fibrosis-inducing agent bleomycin(210). Figure 3.17A shows a t-SNE plot of this data, with cells identified as lipofibroblasts in the original paper clustering strongly together.

In order to compare the human and mouse datasets, firstly, those genes which have a one-to-one mapping between the human and mouse genomes were identified using the BioMart database. 16214 genes out of 27931 mouse and 33694 human were identified as having a 1:1 human:mouse gene mapping – that is, there is only one human gene homologue associated with a particular mouse gene, and vice versa. Although this does remove a lot of the data available in the dataset, it allows a more robust comparison between the two datasets. Clustering both datasets using only these 16214 genes, shown in Figure 3.17, shows that these genes are sufficient to preserve the clusters and shape of the t-SNE dimensional reduction plot, suggesting that the major variation in the dataset is preserved with these genes.

In order to compare the two genesets, the FindMarkers() Seurat function was used to identify marker genes for these cell types compared to the other fibroblasts in both datasets. Figure 3.18 shows the correlation between log FC values in marker genes common between both cell types. It is clear that, although there is a slight positive correlation between the marker genes associated with both cell types, there is a large subset of genes which does not follow this trend, being upregulated in PLIN2-positive cells and downregulated in mouse lipofibroblasts, and vice-versa. Importantly, PPARγ is not among the commonly upregulated genes, although PLIN2 is highly upregulated in both cell types. As PPARγ is one of the most important canonical lipofibroblast markers, its absence is notable(54).

It is therefore hard to conclude from these data that the PLIN2 positive fibroblasts in the IPF data are human lipofibroblasts. At best, they share a common marker gene, PLIN2, with canonical lipofibroblasts, and may share some functional role in maintaining the lung epithelium. However, it is plausible that these cells still act to preserve epithelial functionality in a manner analogous to lipofibroblasts. Given their expression of genes associated with the oxidative stress response, as well as their expression of PLIN2, they may still be important for reducing the impact of profibrotic signalling pathways.

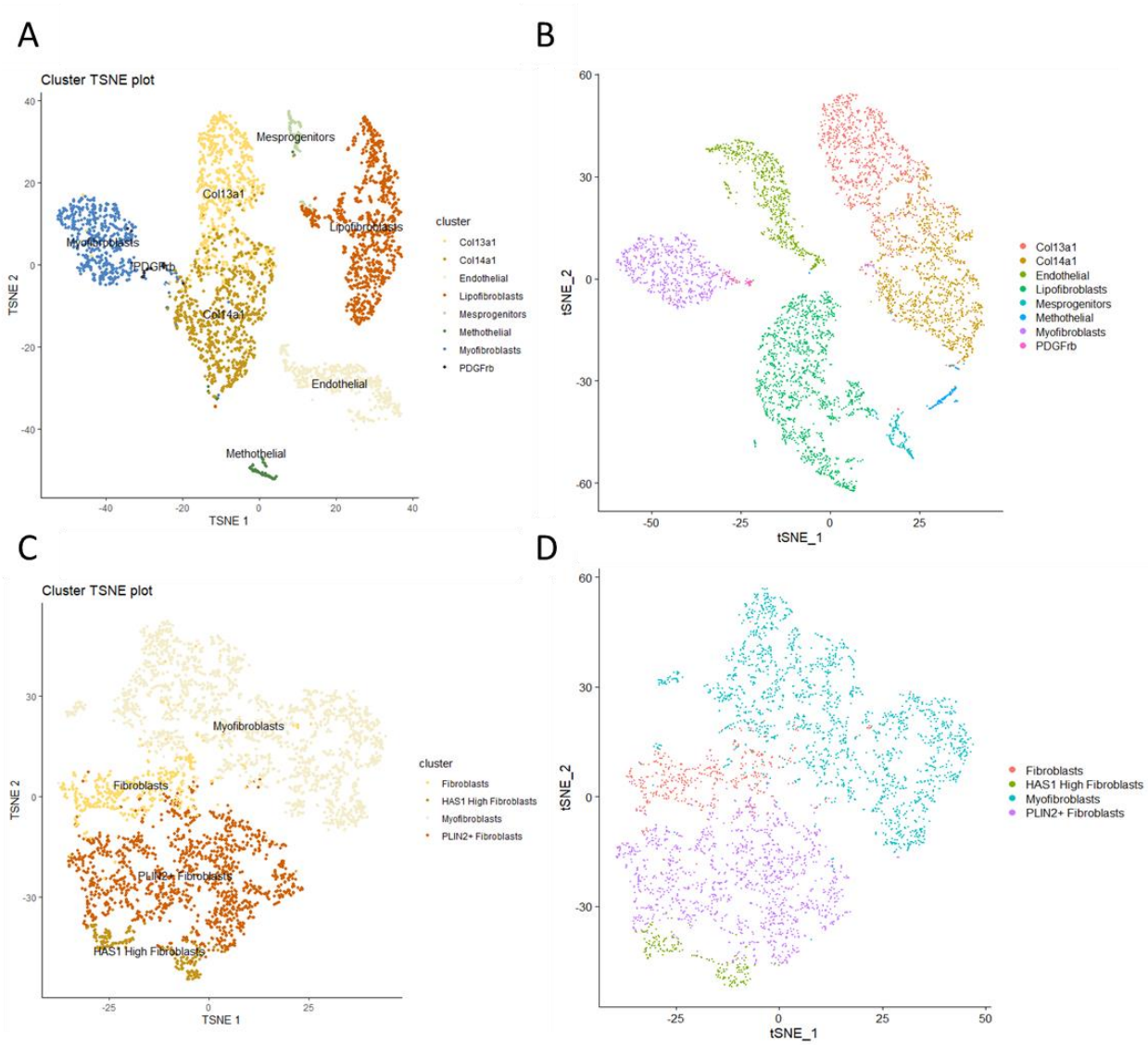


Figure 3.17: TSNE plots showing distribution and clustering of different types of mesenchymal cells in human and mouse datasets. A. and C. Clustering of Jiang and Banovich/Kropski datasets using all genes, showing different fibroblast subtypes identified in their respective publications. B. and D. Clustering of the same datasets using only 16,214 genes with a 1:1 mapping between human and mouse genes.

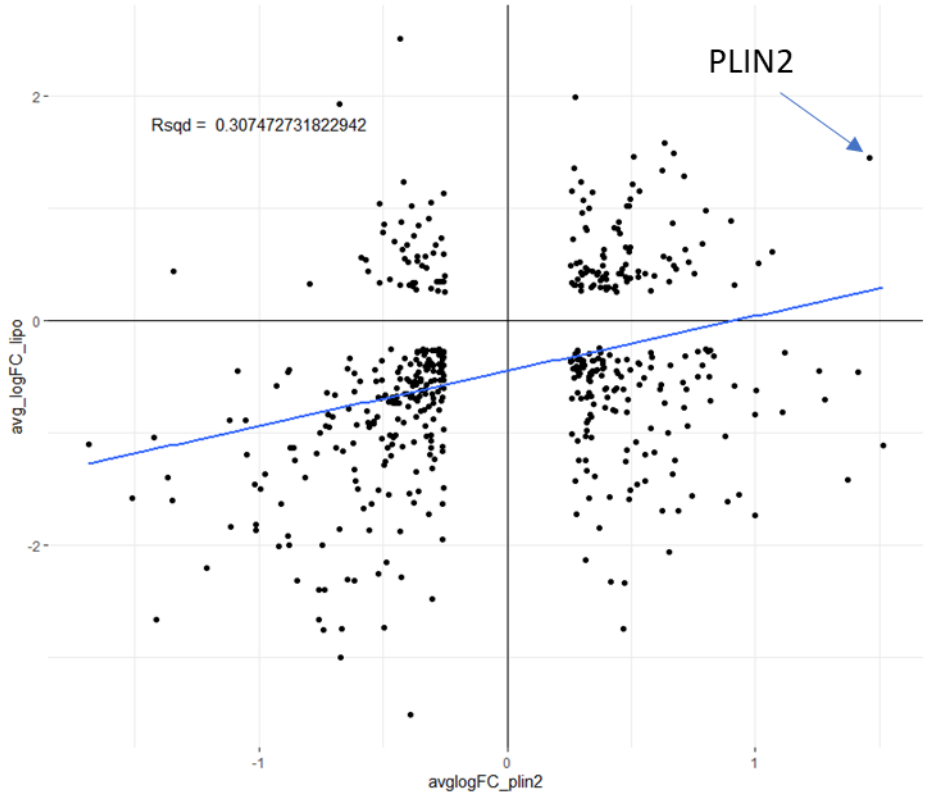


Figure 3.18: Correlation between log_2 fold change of common marker genes found in both human PLIN2-positive fibroblasts and mouse lipofibroblasts. Marker genes identified using Wilcoxon rank sum test on normalised single-cell data using FindMarkers() function in Seurat R package.

3.20 Pseudotime analysis suggests that PLIN2-positive fibroblasts may be a precursor to HAS1-high fibroblasts

It is evident from the t-SNE plot distribution of fibroblasts that the PLIN2 fibroblasts form a continuum with the HAS1 high fibroblasts, with no clearly demarcated clustering, suggesting that they are more similar to each other than they are to other mesenchymal cells in the dataset. Therefore, it is possible to hypothesise that pro-fibrotic, senescent, inflammatory HAS1-high cells are descended from PLIN2 positive cells. Although t-SNE is a useful heuristic to identify common groupings of cells, it does not preserve intra-cluster differences particularly well, and cannot be used for lineage-tracing experiments. In order to investigate whether PLIN2 positive cells are precursors to HAS1-high cells, The R package Monocle's pseudotime analysis functionality was used. This calculates eigengenes which maximise the variance across a distribution – in this case across the fibroblasts clusters, allowing transitions from one state to another to be

identified(180). Instead of clustering cells into different types, it allows smooth variations from one cell type to another to be identified. Firstly, the data are dimensionally reduced using the UMAP algorithm then a graph is calculated which spans individual clusters on the UMAP plot. A user-specified node of this graph is then used as the starting point to identify variation across a cluster, and identify potential cell lineages. This algorithm is most robust when time-series data are available, which is not the case for this dataset, but it can provide insight into the lineages of different cell types.

To test this methodology, epithelial cells from the Banovich/Kropski single-cell RNAseq dataset were subjected to pseudotime analysis. As these have a well-known developmental progression, from alveolar type II to transitional AT2 to Alveolar type I cells, they provided a useful test-case for the pseudotime analysis(211). Figure 3.19A shows a cluster plot for these cells, showing the different epithelial cell types clustering together on a UMAP plot. Figure 3.19B shows the pseudotime scores of individual cells on a colour gradient. The development of ATII cells to transitional ATII, and then branching into ATI cells as well as an IPF specific KRT5-/KRT17+ group of epithelial cells, suggesting this methodology is applicable to this dataset.

Figure 3.19D shows the pseudotime analysis for the IPF fibroblasts, showing a clear continuum of genetic variance from PLIN2-positive fibroblasts to HAS1-high fibroblasts, which cluster separately from the other fibroblast types in this dataset. This suggests that HAS1-high cells constitute a senescent, pro-fibrotic subpopulation of the PLIN2-positive fibroblasts, which instead of contributing to epithelial maintenance, are pathogenic. The presence of PLIN2-positive fibroblast populations outside of the main group – something that was not observed in t-SNE plots, may suggest that there are subpopulations of PLIN2+ fibroblasts within the larger group. However, they cluster together on t-SNE plots, and they were assigned as a single cluster in the original publication, so they were treated as such for single-cell data analysis. Figure 3.19D also shows a clear gradient of pseudotime scores from fibroblasts to myofibroblasts. This suggests that myofibroblasts in lung tissue differentiate from normal fibroblasts.

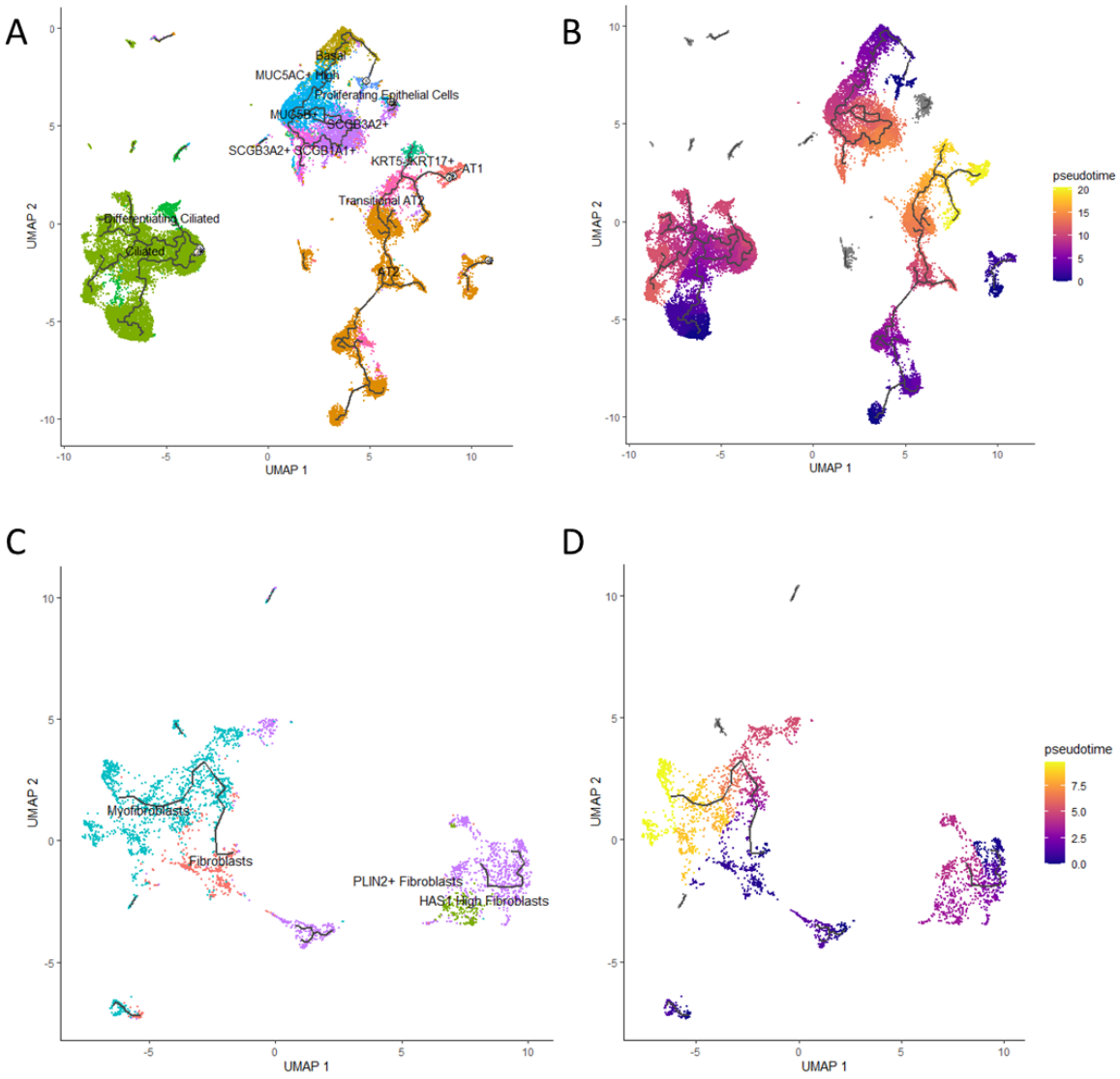


Figure 3.19: Pseudotime analysis of alveolar cells and fibroblasts in Banovich/Kropski dataset. A. and B. UMAP plots of alveolar cells labelled by cell type (A.) and Pseudotime score (B). C. and D. UMAP plots of fibroblasts labelled by cell type and pseudotime score. Pseudotime scores were calculated using the Monocle3 R package and nodes were manually selected to best represent the gradient in the pseudotime scores across clusters.

3.21 WISP1 is expressed in fibroblasts, and particularly in myofibroblasts and HAS1 high fibroblasts

Having characterised the different kinds of mesenchymal cells present in this dataset, identifying the cells expressing WISP1 allows the transcriptional context of WISP1-expressing cells to be defined, providing insight in the functional roles of WISP1 in IPF.

Figure 3.20 shows the expression pattern of the WISP1 gene in the whole dataset, showing a distinct pattern of gene expression localised to particular cell types. WISP-1 is a relatively lowly expressed gene in the dataset as a whole but has higher expression levels in myofibroblasts and HAS1 high fibroblasts. These two cell types were also the most well-represented mesenchymal cells in the fibroblast foci laser capture samples according to the CibersortX cell sorting analysis, verifying the localisation of expression of WISP1 to the fibroblast foci identified in the laser-capture RNA-seq data above. Notably, within these myofibroblasts, the WISP1 expressing cells cluster together, suggesting that these cells have some shared gene expression features even within the larger myofibroblast umbrella. WISP-1 expressing cells are nearly all IPF-derived in this dataset.

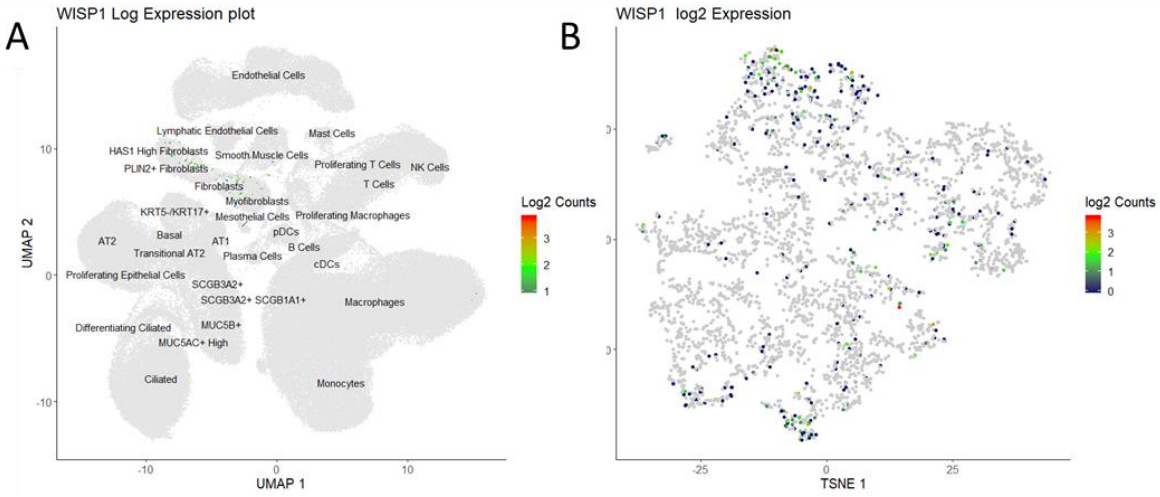


Figure 3.20: Expression pattern of WISP1 in Banovich/Kropski single cell RNA-seq dataset. A. UMAP plot of all cells in the dataset (see Figure 3.8) showing the expression pattern of WISP1. B. TSNE plot of all fibroblasts in the dataset (see Figure 3.13) showing the expression pattern of WISP1. Point colour represents $\log_2$(Counts), with light grey corresponding to 0 counts in the dataset.

3.22 WISP1 expressing myofibroblasts have a pro-fibrotic signal associated with them.

To identify the expression context of WISP1 expressing cells, a gene signature associated with WISP1-expressing cells was identified. For this, the FindMarkers() function was used to compare cells in a particular cluster which expressed WISP1 to those which did not. This provides context for how WISP1 expressing cells behave in IPF. To control for differences in cell type, only cells of the same type (i.e. WISP1 expressing myofibroblasts with non-WISP1 expressing myofibroblasts) were compared.

Table 3.13 shows the top 30 marker genes associated with WISP1 expressing myofibroblasts, while Table 3.14 shows selected significantly upregulated marker genes. There is a very clear enrichment for genes previously associated with fibrosis, with many collagen genes upregulated in these cells. Other enriched genes include SPARC, a gene previously associated with suppression of apoptosis in IPF fibroblasts(195), POSTN, encoding the periostin protein which is associated with epithelial-mesenchymal transition and wound-healing; it has previously been identified as being upregulated in IPF, and suggested to play a role in airway remodelling and fibrosis(212). It

has been also identified as a potential diagnostic biomarker for interstitial lung diseases including IPF(213). Table 3.14 shows selected marker genes for these cells which have been shown to be important for IPF pathogenesis. Notably, a number of collagen genes, as well as LOXL2 and PLOD2 genes are upregulated; expression of PLOD2 and LOXL2 have been associated with bone-type crosslinking and an increase in collagen stiffness in IPF – a major contributor to disease progression(65).

The association of WISP1-expressing cells with multiple known fibrosis markers suggests that the WISP1 gene is either upregulated by fibrotic conditions or plays a role in exacerbating fibrosis. Given its known roles in cell proliferation and bone development, there are previously identified mechanisms by which it may contribute to IPF pathogenesis(150,214).

Table 3.13: Top marker genes ordered by adjusted p-value in WISP1-expressing myofibroblasts compared to non-WISP1 expressing myofibroblasts. Markers were identified using a Wilcoxon rank sum test with the Bonferroni multiple test correction using the FindMarkers() function in the Seurat R package.

Gene	P-Value	LogFC	pct.1	pct.2	p_val_adj
WISP1	0	1.27	1	0	0
CTHRC1	3.16E-21	0.85	0.714	0.424	1.06E-16
SULF1	1.52E-20	0.75	0.567	0.261	5.12E-16
GPC1	1.18E-19	0.35	0.438	0.161	3.96E-15
FAM198B	8.39E-19	0.29	0.36	0.118	2.83E-14
KIAA1217	3.64E-18	0.40	0.296	0.087	1.23E-13
INHBA	8.56E-18	0.48	0.463	0.186	2.88E-13
ITGA5	1.12E-17	0.45	0.419	0.162	3.77E-13
POSTN	5.29E-17	0.95	0.66	0.38	1.78E-12
ADAM12	1.54E-16	0.30	0.271	0.079	5.19E-12
IGFBP3	1.38E-15	0.65	0.453	0.194	4.65E-11
RAI14	1.71E-15	0.37	0.399	0.159	5.76E-11
TPM4	3.76E-15	0.42	0.916	0.755	1.27E-10
COL6A1	4.34E-15	0.46	0.966	0.865	1.46E-10
COL1A1	5.13E-15	0.81	0.961	0.884	1.73E-10
PKM	5.99E-15	0.45	0.798	0.599	2.02E-10
SERPINH1	1.80E-14	0.49	0.823	0.634	6.07E-10
GJB2	2.44E-14	0.30	0.177	0.042	8.23E-10
PDPN	2.94E-14	0.39	0.552	0.269	9.92E-10
SPARC	5.25E-14	0.64	0.985	0.956	1.77E-09
TNFRSF12A	1.33E-13	0.42	0.719	0.422	4.47E-09
COL3A1	1.84E-13	0.57	0.956	0.895	6.19E-09
LOXL2	2.65E-13	0.35	0.3	0.113	8.95E-09
COL1A2	3.99E-13	0.60	0.99	0.965	1.34E-08
MXRA5	6.42E-13	0.43	0.695	0.419	2.16E-08
PDLIM4	6.82E-13	0.36	0.453	0.213	2.30E-08
CALU	1.68E-12	0.36	0.857	0.657	5.65E-08
GPX3	2.67E-12	-0.81	0.394	0.62	9.01E-08
COL6A3	4.35E-12	0.43	0.966	0.865	1.47E-07

Table 3.14: Selected significantly differentially expressed marker genes in WISP1-expressing myofibroblasts compared to non WISP1 expressing myofibroblasts. Genes were selected based on t=previous association with IPF and fibrosis.

Gene	p_val	avg_logFC	pct.1	pct.2	p_val_adj
WISP1	0	1.273101	1	0	0
CTHRC1	3.16E-21	0.845183	0.714	0.424	1.06E-16
SULF1	1.52E-20	0.753855	0.567	0.261	5.12E-16
ITGA5	1.12E-17	0.451111	0.419	0.162	3.77E-13
POSTN	5.29E-17	0.94651	0.66	0.38	1.78E-12
COL6A1	4.34E-15	0.459593	0.966	0.865	1.46E-10
COL1A1	5.13E-15	0.812746	0.961	0.884	1.73E-10
SPARC	5.25E-14	0.640537	0.985	0.956	1.77E-09
COL3A1	1.84E-13	0.568699	0.956	0.895	6.19E-09
LOXL2	2.65E-13	0.347202	0.3	0.113	8.95E-09
COL1A2	3.99E-13	0.600883	0.99	0.965	1.34E-08
COL6A3	4.35E-12	0.428552	0.966	0.865	1.47E-07
COL10A1	5.54E-12	0.340785	0.31	0.123	1.87E-07
COL5A1	3.36E-11	0.33535	0.788	0.562	1.13E-06
COL6A2	1.38E-10	0.258749	0.98	0.915	4.65E-06
PLOD2	2.88E-10	0.290818	0.626	0.372	9.70E-06

3.23 There is a correlation between the pattern of WISP-1 expressing fibroblast subtypes and pro-fibrotic gene expression

Having identified a pro-fibrotic signature in WISP1 expressing myofibroblasts, identifying expression patterns of pro-fibrotic genes across all the fibroblast subtypes in the dataset was important, in order to see if the pattern of WISP1 gene expression was correlated with the expression of pro-fibrotic markers.

Figure 3.21 shows this to be the case, known fibrosis markers and extracellular matrix genes such as COL1A1 show a very clear expression pattern that mirrors the expression of WISP1, high in a

subset of myofibroblasts, and in HAS1-high cells. The expression pattern of WISP1 is therefore reflective of a more general expression pattern in pro-fibrotic genes, with HAS1-high cells and a subset of myfibroblasts expressing high levels of pro-fibrotic genes.

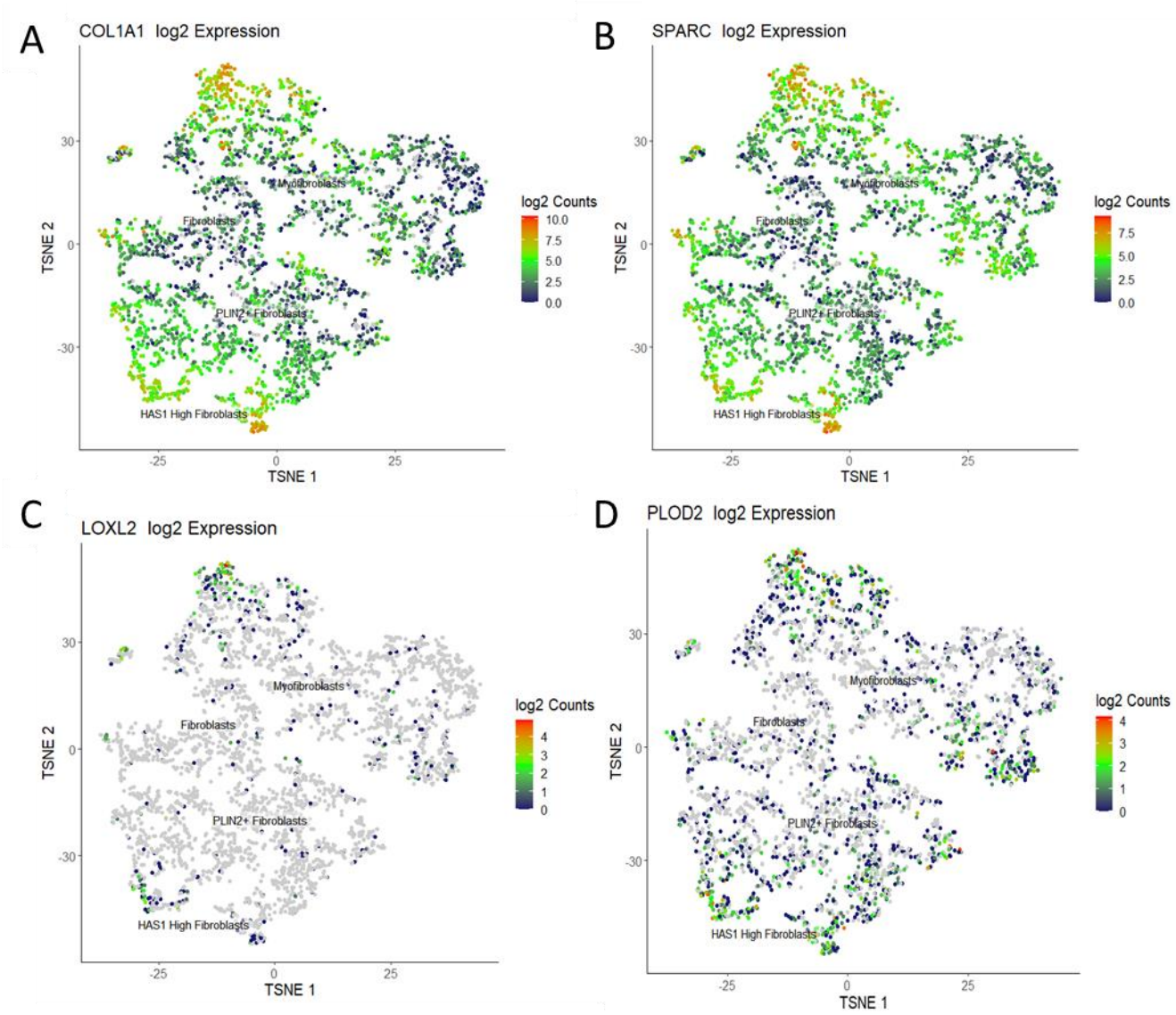


Figure 3.21: Expression pattern of specific IPF-associated genes in fibroblasts in the Banovich/Kropski single cell RNA-seq dataset. Colour represents $\log_2$(Counts). A. COL1A1. B. SPARC. C. LOXL2. D. PLOD2.

3.24 Pro fibrotic gene expression is associated with TGF-β signalling and a HIF activation gene signature

Having identified a correlation with WISP1 expression and pro-fibrotic gene expression, identifying signalling processes which drive expression of these genes was desirable. Gene set variance analysis (GSVA) was used to identify processes which drive WISP1 expression. This calculates an enrichment score based on a ranked gene list for individual samples, in a manner analogous to gene set enrichment analysis. These enrichment scores can be calculated for individual cells in a single-cell RNA-seq dataset and mapped onto a dimensional reduction plot. Three different gene sets were used, representing different signalling processes known to be important for IPF: TGF-β signalling, Wnt signalling, and hypoxia.

The pattern of GSVA enrichment scores was compared to the previously identified WISP-1 and profibrotic gene expression pattern shown in Figure 3.21; a strongly enriched myofibroblast subpopulation and enrichment in HAS-1 high fibroblasts. Figure 3.22 shows GSVA enrichment scores for the fibroblasts in the Banovich/Kropski dataset. Notably, there is negative enrichment of Wnt signalling genes in these cells, with all four fibroblast types in the dataset showing a negative enrichment score, although HAS1-high fibroblasts have the least negative enrichment scores. However, there is enrichment of TGF-β associated genes in HAS1 fibroblasts and myofibroblasts, which show a very similar enrichment pattern to the pro-fibrotic genes identified in Figure 3.21. Given TGF-β's known role as a driver of fibrosis, its correlation with pro-fibrotic gene expression makes sense in the context of IPF. Finally, there is a marked upregulation of HIF pathway-associated genes in the PLIN2+ and HAS1 high fibroblasts. HIF signaling has been reported in lungs and fibroblasts from IPF patients, and HIF activation has been identified as driving expression of collagen crosslinking enzymes important for promoting stiffer, bone-type collagen in IPF(95,96,215). The enrichment of HIF pathway in HAS1-high fibroblasts is particularly interesting given that these cells are major expressors of WISP1.

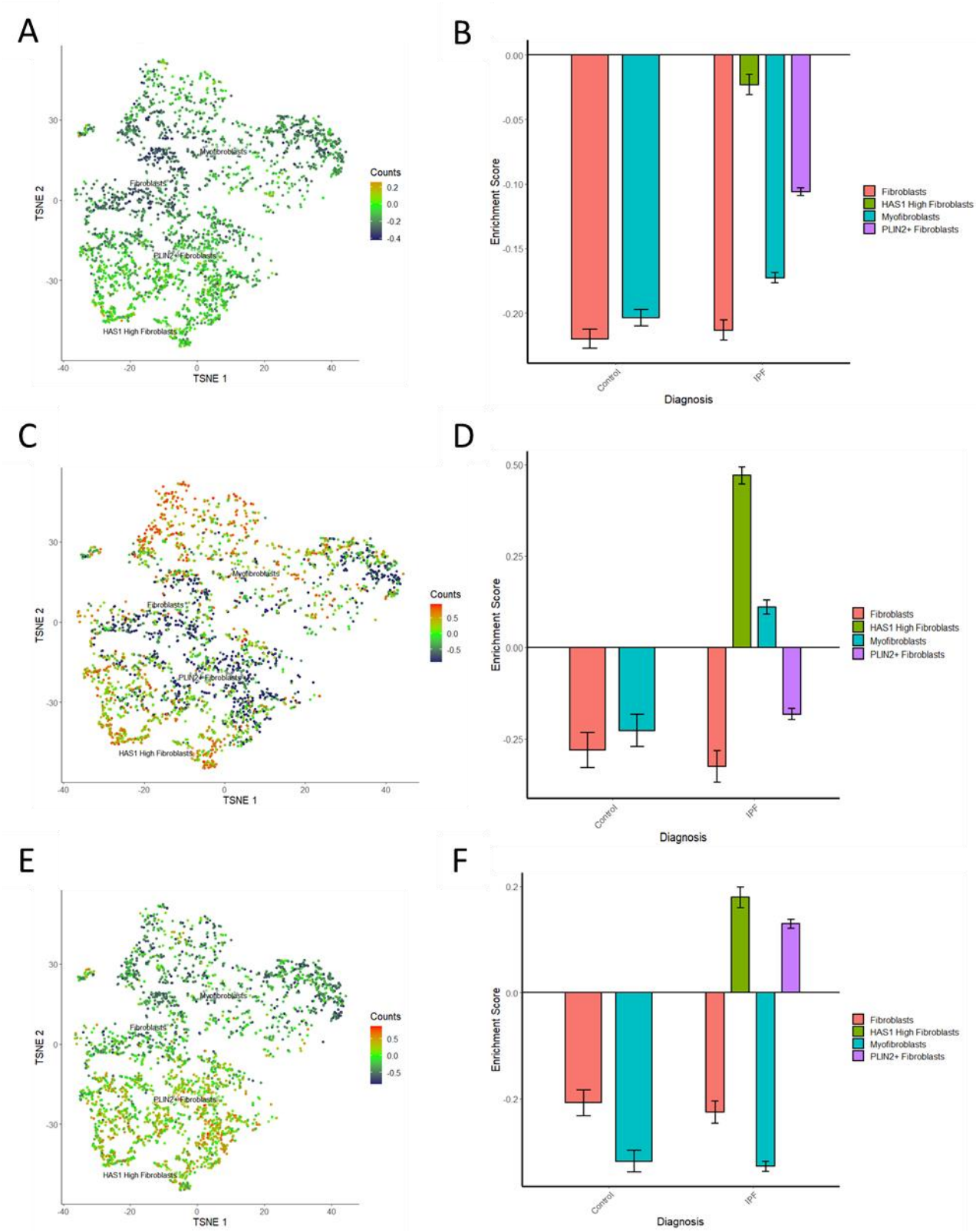


Figure 3.22: t-SNE plots and bar charts showing gene set variance analysis (GSVA) enrichment scores of fibroblasts in the Banovich/Kropski dataset. A and B: Wnt signalling. C and D: TGF-β signalling. E and F: HIF signalling.

3.25 Chapter discussion

This chapter has focussed on identifying gene signatures and patterns associated with IPF, both at a bulk tissue level and using laser-capture RNA-seq and single cell RNA-seq data to identify gene signatures associated with particular cell types and areas of IPF tissue. This has allowed characterisation of the cell types associated with WISP1 expression, and the gene expression signatures associated with those cells expressing WISP1.

This study has identified WISP1 as being largely expressed in the fibroblast foci in IPF, and within this tissue context has identified myofibroblasts and HAS1-high cells as being the principal sources of WISP1 expression in IPF. WISP1 expression is highly correlated to other indicators of fibrosis, including expression of several collagen genes and genes associated with increased collagen stiffness and bone-type crosslinking, as well as other genes that have previously been associated with the disease progression of IPF(65,195).

This study has also more generally identified gene signatures associated with IPF, and with particular tissue types in IPF. A common differential expression signature across datasets confirms the previously identified Wnt and TGF-β signalling networks as being dysregulated in the disease. Characterisation of the different cell types present in IPF has allowed identification of gene expression signatures associated with those cell types, and CibersortX *in silico* cell sorting has allowed localisation of those cell types to specific areas of IPF tissue.

The results presented in this chapter, then, give a comprehensive overview of the gene expression environment in IPF, the localisation of specific cell types within the disease, the localisation, cell type-specific expression and transcriptomic environment of WISP1.

The datasets used for this study each have their individual limitations, some of which are mitigated by the other datasets, some of which are not. Firstly, the bulk RNA-seq and microarray datasets do not adequately capture the microscopic heterogeneity of IPF tissue. Lung tissue in individuals with IPF is a mixture of areas of relatively normal-appearing alveolar epithelium, cysts lined with bronchial epithelial cells, and fibroblast foci. Taking areas of bulk tissue in IPF is therefore a mixture of lots of different, distinct, tissue types, each with their characteristic gene expression signatures. Bulk tissue transcriptomics can only reveal general, large-scale trends in changes in gene expression, and provides no tissue-specific information.

The use of laser-capture microdissection RNA-seq data seeks to address this by extracting RNA only from areas of interest – alveolar septae and fibroblast foci. This succeeds in identifying genetic signatures associated with particular tissue types in IPF and allows localisation of

differential gene expression to known histological features in IPF tissue. However, there are still issues with this approach, notably that these histological areas are still highly complex – alveolar septae contain different types of alveolar epithelial cells, immune cells, and endothelial cells, for example. Although this issue is less acute as with the bulk expression data, it is still present. The other issue is one of adequately controlling for differences in cellular composition between the three different tissue areas sampled. While comparison of gene expression in alveolar septae in IPF and healthy control tissue will reflect both changes in the cellular makeup of this tissue and the gene expression signatures of the tissue brought on by disease, it compares like-for-like tissue, with the roles of alveolar septae the same in IPF and healthy individuals. However, there is no such analogue for fibroblast foci in the control tissue – healthy lung tissue does not have large, multicellular fibroblast aggregates in the same way that IPF lung tissue does. Thus, comparison between the fibroblast foci and healthy lung tissue is as much a comparison of fibroblast gene signatures compared to alveolar epithelial gene signatures as it is for IPF to control tissue.

Finally, single cell RNA-seq data has significant issues associated with it. The first, inherent in the single-cell RNA-seq technology itself, is a large amount of missing data, in the form of genes whose count value is 0 in a given cell, and by extension the limited resolution of the data. This may capture actual gene expression patterns – a particular gene may not be expressed in a cell – but it may also be a result of the limited sensitivity of single-cell RNA-seq methods – the mRNA may have escaped the tagging procedure or been degraded before being sequenced. Therefore, for low-expressing genes, including WISP1, some cells which do express WISP1 may not identify it. A second issue is that single cell RNA-seq data provides no tissue localisation data, unlike the LCMD data described above. Although it can identify many different cell types, the disaggregation procedure used in preparation of a single cell suspension removes all tissue localisation data. This is one motivation for using *in silico* FACS algorithms such as CibersortX to identify where in the lung a particular cell type is likely to be found.

However, all three types of data used in this study still provide useful insight. The bulk RNA-seq data allows the broad gene signatures associated with fibrosis to be identified, and the use of two separate datasets allowed a robust, IPF-specific gene expression signature to be identified. The laser capture microdissection data provides tissue localisation data and allows tissue-specific gene signatures to be identified. The single cell data allows gene expression signatures associated with individual cell types to be identified and allows classification of many individual cell types. It can then be combined with the laser capture microdissection RNA-seq data using CibersortX to allow localisation of specific cell types in IPF tissue.

There is also commonality between the differentially expressed genes identified in the different datasets used in the study. Asporin (ASPN), a gene associated with cartilage development and TGF-β regulation, is upregulated in IPF in both bulk RNA-seq datasets, in IPF alveolar septae compared to control alveolar septae, in fibroblast foci, and is a marker gene for myofibroblasts in the single cell-RNA-seq dataset. ASPN upregulation in IPF is likely a consequence of increased TGF-β signalling in the disease, as evidenced by its upregulation in myofibroblasts, known to be driven by TGF-β signalling(47). Notably, asporin is important in downregulating TGF-β signalling and extracellular matrix development in articular cartilage, suggesting it might be important as a negative feedback indicator for the increased TGF-β signalling identified in IPF(188).

Another gene family which is upregulated in multiple different datasets is the secreted frizzled receptor proteins, or SFRPs. These are Wnt antagonists, and may reflect the role of Wnt signalling in IPF as their expression can be induced by specific Wnt ligands in IPF (216).

There are also many genes and gene ontology terms which are common across multiple datasets, lending confidence that, although the datasets and methodologies are different, they are all still capable of identifying fibrotic gene expression signatures. Many collagen genes are upregulated in both bulk datasets and the LCMD data, and gene ontology terms associated with extracellular matrix development are associated with the common differentially expressed genes in the bulk datasets, as well as the fibroblast foci and IPF alveolar septae. Collagen genes and extracellular matrix development gene ontology terms are also strongly associated with myofibroblasts in the single-cell RNA-seq data, which reflects their known role as ECM producers in fibrosis.

The results shown here also show agreement with previous studies which have identified an increase in collagen stiffness as being important for IPF pathogenesis (65,217), and has attributed that increased collagen stiffness to an increase in bone-type collagen crosslinking, driven by increased expression of the collagen crosslinking enzymes LOXL2 and PLOD2 (65). Several gene ontology terms associated with bone development, including skeletal system development and replacement ossification, are upregulated in fibroblast foci in the LCMD data, while LOXL2 and PLOD2 expression is strongly correlated to WISP1 expression in the single cell dataset, and both genes are upregulated in myofibroblasts and HAS1-high fibroblasts.

Further support for the validity of this study comes from the CibersortX *in silico* FACs results. These identify the different cellular compositions in the different LCMD RNA-seq samples. Several trends from this analysis match known knowledge about IPF. Firstly, there is a very strong fibroblast signature associated with the fibroblast foci samples, with myofibroblasts being a major part of that signature. This agrees with the known cellular composition and function of fibroblast

foci – aggregates of actively matrix-secreting myofibroblasts. Secondly, there is a strong increase in the proportion of endothelial and ciliated cells in IPF alveolar septae compared to control. This is consistent with known remodelling changes found in IPF tissue(79) .

Having established that the datasets identified in this chapter can be used to recapitulate previously known things about IPF, it can be used to form new insights into how the disease works, including the expression patterns and transcriptomic context of WISP1.

According to the laser capture microdissection data, WISP1 expression appears strongly localised to fibroblast foci in IPF, with expression in IPF and control alveolar septae being substantially lower. This is confirmed with the single cell RNA-seq data, where the only cells where WISP1 is detected are fibroblasts. WISP1 is expressed in myofibroblasts and HAS1-high fibroblasts – both cell types which are strongly represented in the fibroblast foci according to the CibersortX analysis. This runs counter to some previous studies which have identified WISP1 as being expressed by alveolar cells in IPF (191,192). This discrepancy may be due to these studies assessing WISP1 localisation using immunohistochemistry, looking at protein localisation as opposed to expression origin. It is possible that WISP1 binds to receptors on the alveolar epithelium adjacent to fibroblast foci, facilitating epithelial-foci crosstalk. WISP1 was not differentially expressed in either bulk transcriptomic dataset, but is expressed at such a low level in both datasets that any differential expression is unlikely to be identifiable – again highlighting issues with tissue heterogeneity in bulk transcriptomic studies.

It is also notable that WISP1 expression in the single cell dataset is very strongly associated with pro-fibrotic genes, including the collagen crosslinking enzymes LOXL2 and PLOD2, several fibrillar collagen genes including COL1A1, COL6A1 and COL3A1, and SPARC (osteonectin), a gene which has previously been associated with fibroblast survival and proliferation in IPF as well as being important for bone development. This suggests that WISP1 is involved directly in fibrosis, either by driving expression of pro-fibrotic genes or being itself driven by a pro-fibrotic signalling pathway.

The origin and expression patterns of the cell types which express WISP1 – the HAS1 high fibroblasts and myofibroblasts, is also of interest. Pseudotime analysis suggests that HAS1-high cells develop from PLIN2-positive fibroblasts, while normal fibroblasts differentiate into myofibroblasts. This, coupled with the increased expression of secretory factors in the HAS-high fibroblasts, suggests that they are senescent, pro-inflammatory PLIN2 positive fibroblasts. The PLIN2 positive fibroblasts, by contrast, show expression of lipofibroblasts markers, suggesting that these cells may proliferate in IPF as a defence mechanism against disease progression. However,

comparison of the gene expression signatures of mouse lipofibroblasts to these PLIN2 positive fibroblasts does suggest that, while they may share some common marker genes, they cannot be described as lipofibroblasts, especially given their lack of PPARγ signalling. Nonetheless, their expression of genes associated with mitigation of oxidative stress as well as their expression of PLIN2, which codes for a lipid-droplet binding protein characteristic of lipofibroblasts suggests they may still plausibly have some protective effect in mitigating fibrosis(206).

The single cell dataset also allows exploration of the different roles mesenchymal cells may play in IPF. The classic picture of these cells in IPF is that of actively matrix-secreting myofibroblasts which produce the stiffened, fibrotic extracellular matrix characteristic of the disease. The CibersortX analysis demonstrates that myofibroblasts make up a large proportion of the fibroblasts in fibroblast foci. However, the analysis also identified a large percentage of HAS1-high fibroblasts in the fibroblast foci laser capture samples, as well as a smaller number of PLIN2-positive fibroblasts. HAS1-high cells, which express many pro-fibrosis markers as well as indicators of being senescent through their highly secretory and inflammatory gene expression signature, likely contribute to fibrosis, and may be important for maintenance of the extracellular matrix production seen in IPF. The PLIN2 positive cells, meanwhile, express several markers such as metallothioneins, which may provide protection against oxidative stress, and PLIN2 itself – a protein important for epithelial cell maintenance.

Having identified the expression context of WISP1, and the types of fibroblast present in IPF, the questions as to what drives WISP1 expression in IPF remains unclear. There is clear evidence of Wnt and TGF-β signalling in both the laser capture and bulk datasets, but less clear evidence of Wnt signalling being present in the WISP1 producing fibroblasts. There is some correlation between WISP1/fibrotic gene expression and the expression of HIF markers, and better correlation between pro-fibrotic genes and TGF-β gene set enrichment, fitting the known role of TGF-β as a driver of fibrosis, but this will require further investigation.

To conclude, this investigation has identified a core gene signature associated with fibrosis that is common to multiple transcriptomic datasets using many different methodologies. It has confirmed known roles of fibroblast foci in producing extracellular matrix, as well as identifying known processes, such as upregulation of the enzymes which catalyse bone-type collagen crosslinking, which have been shown to be important for fibrosis. This study has identified that WISP1 is more highly expressed in fibroblast foci than alveolar septae and has identified multiple types of foci-resident mesenchymal cells, including classical fibroblasts and myofibroblasts, PLIN2 positive fibroblasts which may have a protective role in IPF, and HAS1-high fibroblasts which are

likely pro-fibrotic, senescent PLIN2-positive fibroblasts. The latter two are principally found in IPF only. Of these cell types, WISP1 is expressed in myofibroblasts and HAS1-high fibroblasts, and WISP1 expression is very strongly correlated with the expression of extracellular matrix components and known pro-fibrotic markers, including the bone-type crosslinking enzymes previously identified as important for pathogenesis. Drivers for WISP1 and these pro-fibrotic markers remain open to investigation, as elements of one of the principal signalling pathways identified in bulk and LCMD datasets, Wnt signalling, do not correlate well with the expression of these markers. However, there is some correlation between these markers and HIF and TGF-β marker genes.

4 Investigating drivers of WISP-1 in lung fibroblasts

4.1 Introduction

Building on the previous chapter which identified mesenchymal cells as the primary site of expression of WISP-1 in IPF, the purpose of this chapter is to identify drivers of WISP1 expression. Classically, WISP1 is a Wnt-induced protein(131,132) – however, attempts to identify a Wnt signature in IPF using the bioinformatic approaches detailed in the previous chapter, were inconclusive. The aims of this chapter are to examine the effects of known drivers of pulmonary fibrosis, such as TGFβ and hypoxia(94,218), as well as Wnt signalling on WISP-1 expression and to confirm the localisation of WISP1 expression to the fibroblast foci in IPF tissue using RNAscope *in-situ* hybridisation.

From its initial discovery, WISP-1 has been associated with Wnt signalling(131,132). Although little evidence was found for a Wnt-signalling signature in IPF in the previous chapter, other studies have identified Wnt signalling as being important in IPF pathogenesis(142,191,192). IPF is also a disease where repair and development pathways such as Wnt signalling are pathologically activated(111,113).

However, other signalling pathways contribute to IPF. The prototypic pro-fibrotic pathway is TGF-β signalling, which is highly upregulated in fibrotic tissue(218). It regulates multiple pro-fibrotic processes, including cellular proliferation, fibroblast activation and ECM remodelling(219). Recent research has also highlighted the role of hypoxia in IPF pathogenesis(95,110,215). Impaired gas exchange caused by interstitial fibrosis leads to an increase in hypoxia signalling in IPF, and this has been implicated in several known pro-fibrotic processes, including collagen crosslinking(96,215). Research detailed in the previous chapter also suggests that hypoxia may be critical in driving the phenotype of HAS1-high cells and profibrotic myofibroblasts which produce the majority of WISP1 mRNA in the single-cell RNAseq dataset.

This chapter investigates drivers of WISP-1 expression in cultured lung fibroblasts using these three known profibrotic pathways, Wnt, TGF-β, and hypoxia. By adding Wnt ligands, TGF-β or prolyl hydroxylase inhibitors (DMOG or IOX2 that mimic hypoxia by stabilising HIF) to cell culture media cells or by growing cells in hypoxic conditions, WISP-1 is demonstrated to be upregulated by hypoxia, and specifically HIF activation. The research presented here also shows HIF activation induces expression of WISP1 splice variants. This chapter also examines the effect of siRNA-based knockdown of WISP-1 in both IPF and normal lung fibroblasts.

4.2 RNAscope *in-situ* hybridisation of IPF tissue shows that WISP1 and WNT5A mRNA localises to fibroblastic foci

The type of cells which express WISP1 in IPF is important both to identify how it affects the disease process, but also in the selection of which cell types to investigate drivers and functions of WISP-1. Research shown in the previous chapter identified WISP1 expression as being largely confined to the fibroblastic foci in IPF tissue. To confirm this site of expression, RNAscope *in-situ* hybridisation was employed using IPF lung tissue and control non-fibrotic healthy lung tissue (N=7 IPF, 3 healthy). In addition to WISP1, probes for WNT3A and WNT5A were also tested. While there are a multitude of different Wnt ligands, Wnt3A and 5A are the prototypical ligands for canonical and non-canonical Wnt signalling respectively(220–222). As WISP1 was initially described as a Wnt-inducible gene, identifying whether Wnt ligands are expressed in the same tissue area as WISP-1 is key in determining whether WISP1 expression is Wnt driven. In addition, tissue was immunostained for the hypoxia-inducible gene product, carbonic anhydrase 9 (CA9).

As shown in Figure 4.1A and B, the WISP1 mRNA signature was found to localise to fibroblastic foci, with cells showing positive staining. This is a representative fibroblastic focus, shown in serial sections with different probes. WISP1 mRNA consistently localised to fibroblastic foci cross multiple donors. No staining for WNT3A mRNA was observed in IPF fibroblast foci (Figure 4.1C and D), while WNT5A mRNA expression was found to localise to fibroblast foci (Figure 4.1E and F). CA9 immunostaining was also evident in fibroblastic foci.

These data confirm that WISP1 is expressed in fibroblast foci, and suggests that a fibroblast cell culture model is the best for investigating drivers of WISP1 expression. This is in contrast with a previous study that suggested ATII cells are the primary site of WISP-1 expression(191,192). However, staining in these studies was not able to be replicated – Commercial antibodies against WISP-1 for IHC were tested by a previous PhD student and did not give consistent patterns of staining. In addition, attempts to block binding with exogenous WISP1 were unsuccessful. More specific antibodies have been generated by some companies, however their use in published work is restricted by the company's own licensing agreements. Development of antibodies in-house was not considered, as the main purpose of tissue localisation studies was to confirm the site of WISP-1 RNA expression, which was achieved using RNAscope. This permitted confirmation that WISP-1 was expressed in fibroblastic foci and a decision to be made to study WISP-1 expression and function in fibroblasts rather than epithelial cells.

Furthermore, the *in-situ* hybridisation studies also suggest that if Wnt signalling is responsible for induction of WISP1 expression in IPF, it is likely to be via non-canonical Wnt signalling. Upregulation of WNT5A mRNA in fibroblastic foci may reflect the known role of Wnt5A in regulating fibroblast proliferation and apoptosis resistance in pulmonary fibrosis(223).

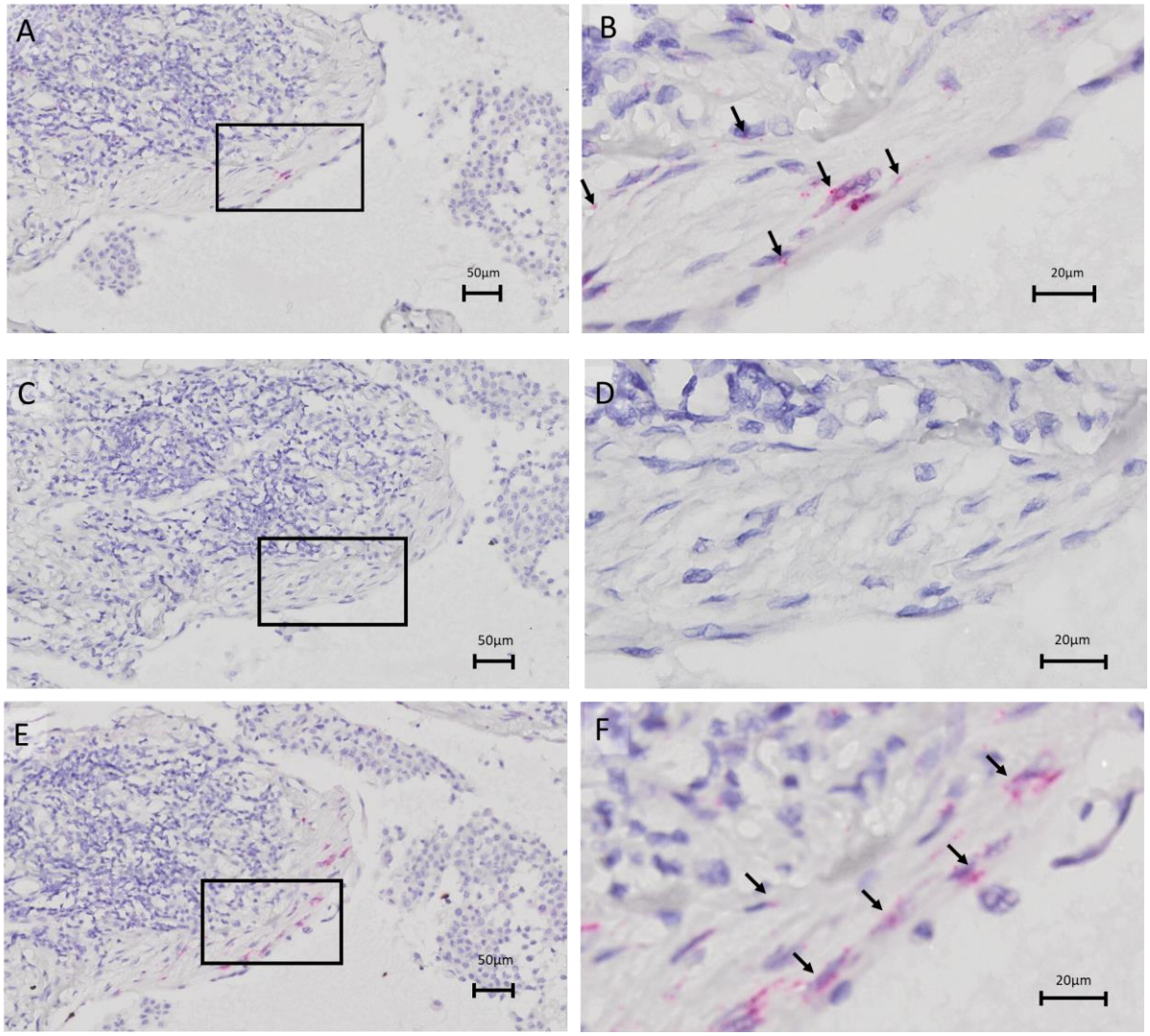


Figure 4.1: RNAScope in-situ hybridisation of mRNA coding for WISP1, WNT3A and WNT5A in a fibroblast focus. Images are of serial sections taken from video-assisted thoracoscopic (VATS) lung tissue biopsy. Left-hand images are with a 10x objective, right-hand images are with a 40x objective. Black boxes denote magnified areas in right-hand images. Red dots are staining for mRNA molecules. Arrows show representative areas of staining. Blue colour is haematoxylin counterstaining. A. and B. WISP1 mRNA expression. C. and D. WNT3A mRNA expression in IPF tissue. E. and F. WNT5A expression in IPF tissue.

4.3 RNAscope *in-situ* hybridisation of IPF alveolar septae shows no WISP1 or WNT3A expression, but WNT5A mRNA is expressed.

Figure 4.2 shows representative RNAscope results from IPF alveolar septae – areas of alveolar tissue from fibrotic lung, but which do not contain fibroblastic foci. There is no evidence of WISP1 or WNT3A mRNA *in-situ* hybridisation in alveolar septae. However, WNT5A mRNA was detected, albeit at a lower level than in fibroblastic foci. This corroborates a previous study which identified Wnt5A protein as being expressed in both fibroblastic foci and alveolar epithelium in IPF tissue(224).

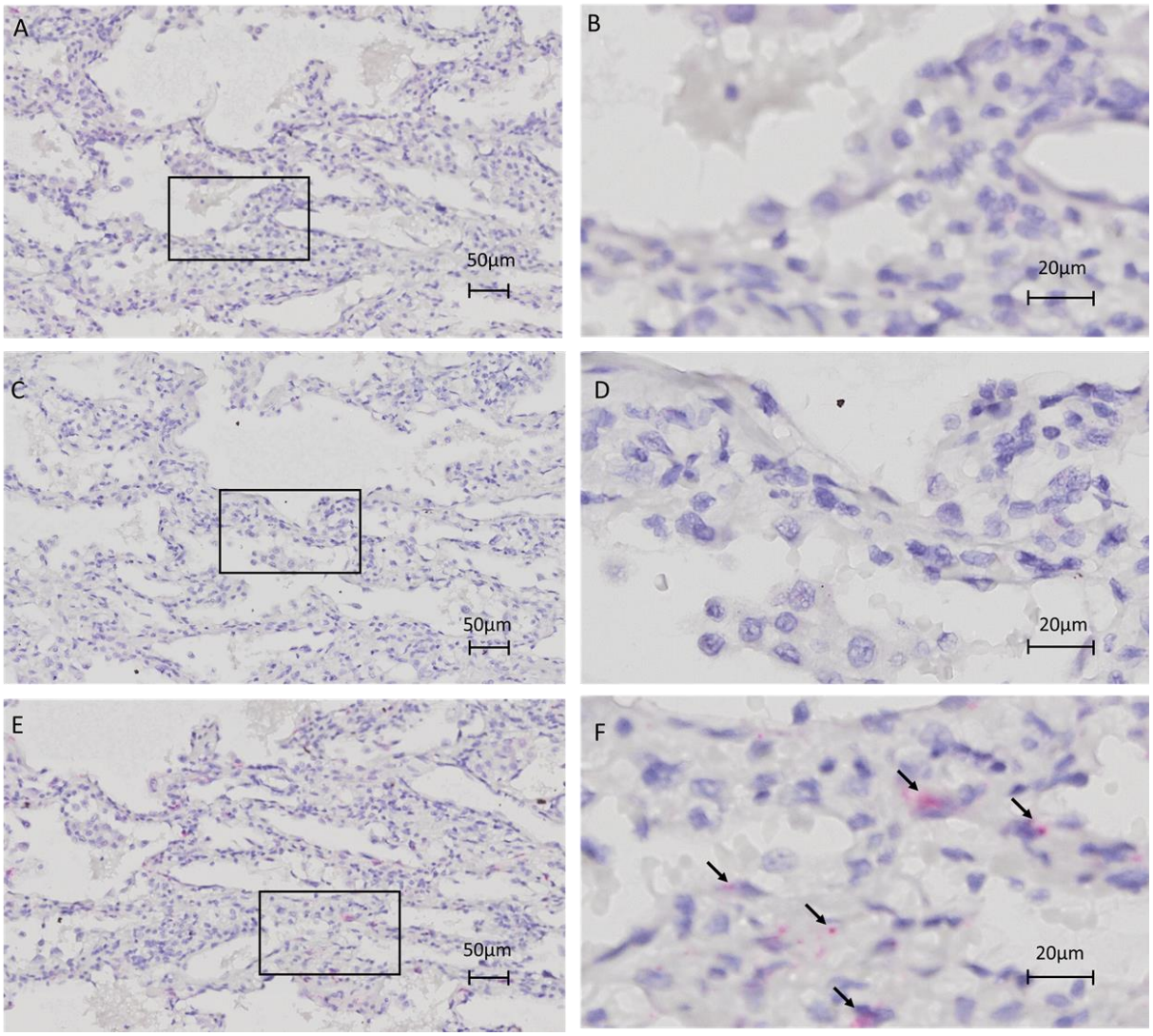


Figure 4.2: RNAScope in-situ hybridisation of mRNA coding for WISP1, WNT3A and WNT5A in IPF alveolar septae. Images are of serial sections taken from VATS lung tissue biopsy. Left-hand images are with a 10x objective, right-hand images are with a 40x objective. Black boxes denote magnified areas in right-hand images. Red dots are staining for mRNA molecules. Arrows show representative areas of staining. Blue colour is haematoxylin counterstaining. A. and B. WISP1 mRNA expression. C. and D. WNT3A mRNA expression in IPF tissue. E. and F. WNT5A expression in IPF tissue.

4.4 RNAscope *in-situ* hybridisation of nonfibrotic lung tissue shows no WISP1, WNT3A or WNT5A expression.

Figure 4.3 shows the representative images of RNAscope *in-situ* hybridsation of WISP1, WNT5A and WNT3A mRNA in lung tissue taken from normal-appearing lung surrounding a cancer resection. No detectable staining for any of these genes was identified in healthy lung tissue. Wnt signalling is a hallmark of developing tissue, or a wound-healing response, as well as IPF, so lack of Wnt ligand expression in healthy lung tissue is unsurprising. WISP-1 is also important for cellular proliferation, differentiation and survival, processes which are inactive in healthy, uninjured lung tissue(140,150,214).

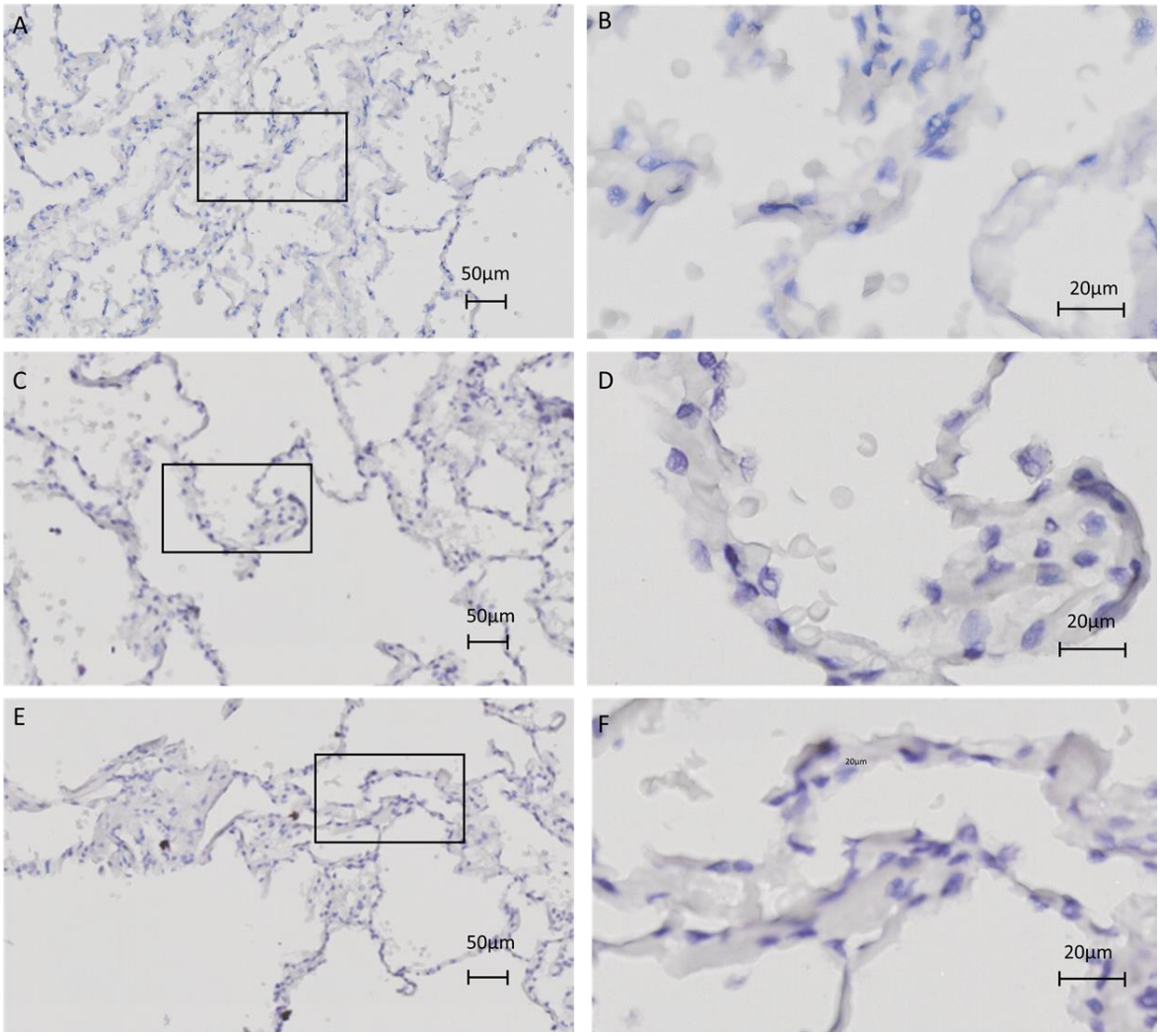


Figure 4.3: RNAScope in-situ hybridisation of mRNA coding for WISP1, WNT3A and WNT5A in healthy alveolar septae. Images are serial sections taken from normal appearing lung tissue surrounding a cancer resection. Left-hand images are with a 10x objective, right-hand images are with a 40x objective. Black boxes denote magnified areas in right-hand images. Red dots are staining for mRNA molecules. Arrows show representative areas of staining. Blue colour is haematoxylin counterstaining. A. and B. WISP1 mRNA expression. C. and D. WNT3A mRNA expression in IPF tissue. E. and F. WNT5A expression in normal lung tissue.

4.5 Immunohistochemical staining reveals the presence of CA9, a HIF responsive protein, in IPF tissue.

Having confirmed the expression of WISP1 and WNT5A in IPF tissue using RNAscope *in-situ* hybridisation, evidence of HIF signalling was sought. Figure 4.4 shows H&E and Carbonic andhydrase 9 (CA9) staining of serial sections of IPF lung, including a fibroblastic focus. The CA9 gene is induced by HIF signalling, so it acts as a marker for HIF signalling activity in this tissue. Staining and imaging was performed by Dr Christopher Brereton. Brown staining corresponds to CA9 protein. There is strong staining for CA9 in this tissue, indicating that HIF signalling is active. CA9 staining is present in fibroblastic foci, suggesting that HIF signalling pathways are active in fibroblastic foci.

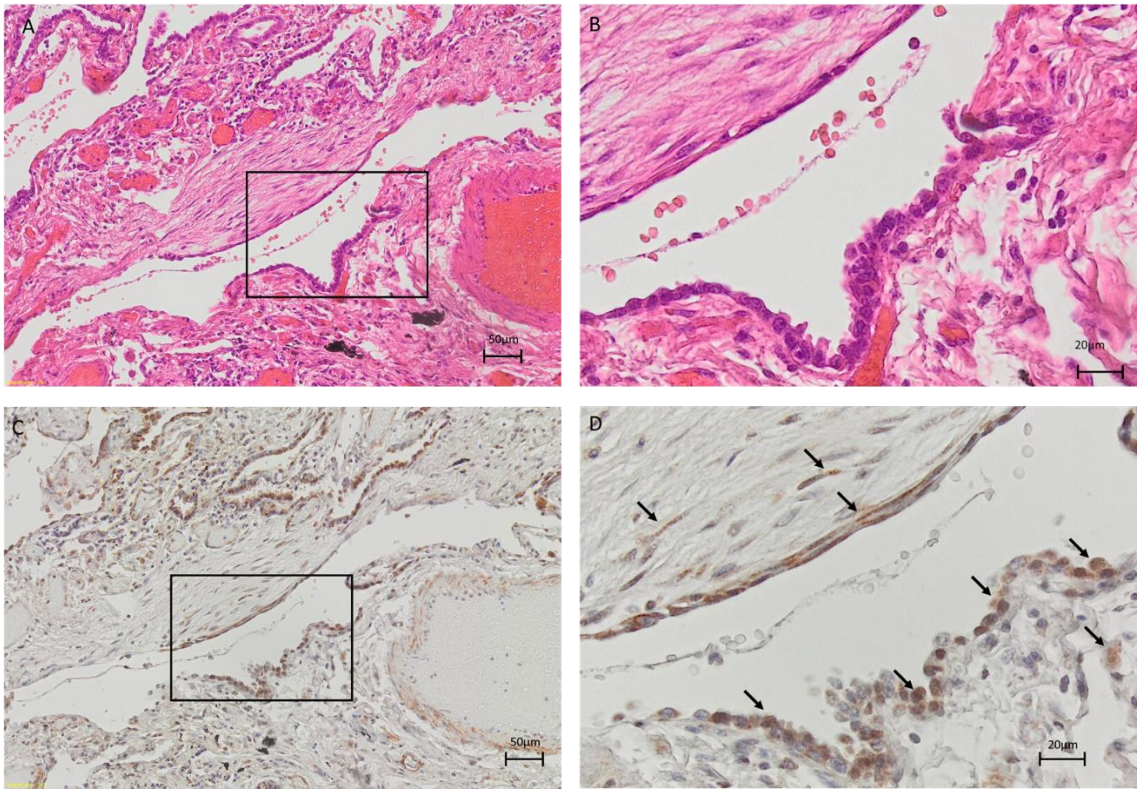


Figure 4.4: haematoxylin and eosin (H&E) and CA9 staining of IPF tissue. Images are of serial sections taken from VATS lung tissue biopsy. Left-hand images are enlarged versions of the black box in right-hand images. Images are courtesy of Dr Chris Brereton. Top. H&E staining of IPF lung. Haematoxylin stains nuclei, eosin stains for extracellular matrix. Bottom. CA9 staining of IPF tissue. Brown dots show CA9 as a marker of HIF activity.

4.6 WISP1 expression in cultured fibroblasts is driven by hypoxia signalling.

In order to investigate drivers of WISP1 expression, initial experiments were performed using primary fibroblasts treated for 24, 48 and 72 hours with epidermal growth factor, TGFβ1, Dimethoxyallylglycine (DMOG, an inhibitor of the oxygen-sensing prolyl hydroxylase enzyme which targets HIF molecules for degradation, thus mimicking the effect of hypoxia), and Wnt3A and Wnt5A, prototypical canonical and non-canonical Wnt signalling ligands. (Figure 4.5). At each time point, RNA was extracted, and WISP-1 expression measured by RT-qPCR. At 24 and 48hrs, WISP1 mRNA levels were highest in cells treated with TGFβ1, although this change was not statistically significant. By 72 hours, however, the hypoxia analogue DMOG caused some increase in WISP1 expression, but again this change was not statistically significant.

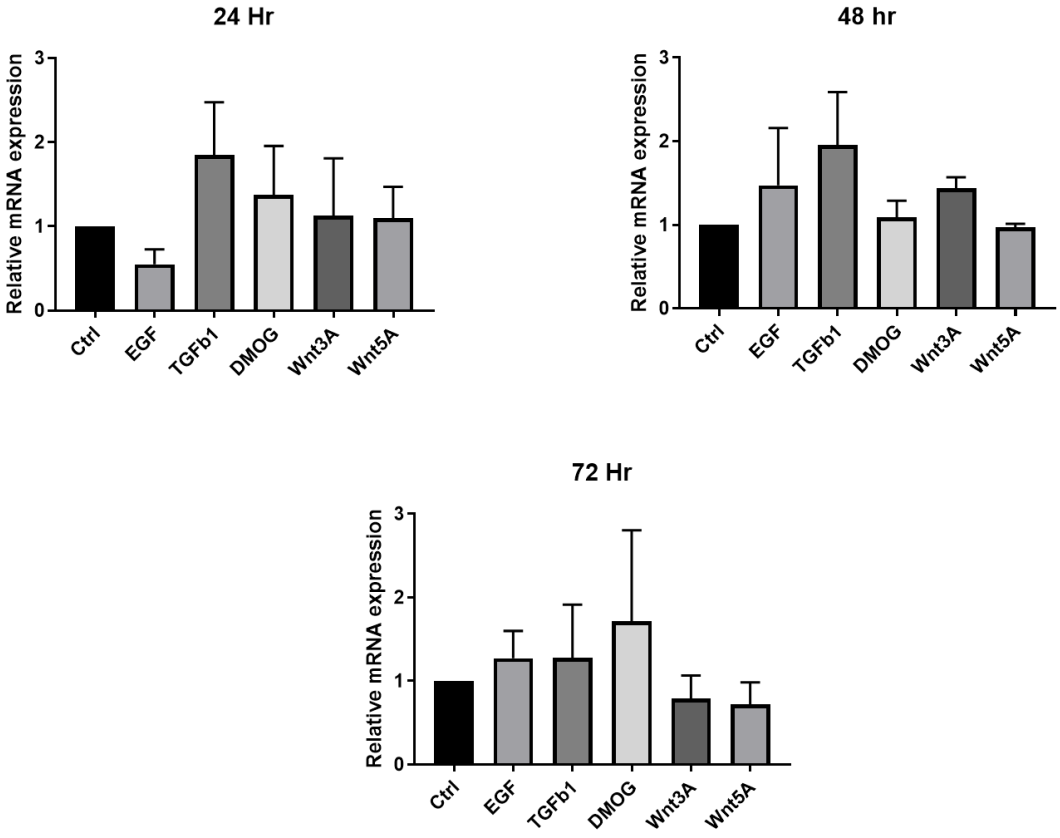


Figure 4.5: 72-hour time course of primary normal fibroblasts treated with several fibrotic mediators. Cells were treated with epidermal growth factor (EGF (10ng/ml), TGF-beta (10ng/ml), DMOG (1mM), Wnt3a and Wnt5A (both 100ng/ml). cDNAs were kindly provided by Dr Christopher Brereton. Treatment effect on cells was confirmed by western blot of phospho-ERK,

SMAD2/3, HIF1α, and β-catenin. WISP1 expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC and A2 housekeeping gene expression C_t values and timepoint control C_t value. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. N = 3 for all treatments. Error bars show standard deviation. No significant difference was found between groups using a one-way ANOVA.

4.7 WISP1 expression is driven by hypoxia in an alveolar type II-derived cell line and MRC-5 fibroblasts.

In order to further investigate the expression of WISP1, a cell line derived from ATII cells, ATII ER: KRASV12 cells and MRC-5 fibroblasts were grown under normoxic and hypoxic conditions, with and without TGFβ, as these were the two largest drivers of WISP1 identified in cultured fibroblasts. ATII cells were tested given previous research demonstrating WISP1 induction in alveolar epithelial cells(191).

In ATII cells, WISP1 was strongly upregulated by a combination of hypoxia and TGFβ, although either treatment alone did not significantly increase WISP1 expression (Figure 4.6A). In MRC-5 fibroblasts, the expression of WISP1 was significantly induced by hypoxia, but not by TGFβ and, in contrast with the ATII cells, no synergistic effect of hypoxia and TGFβ was observed (Figure 4.6B). Most importantly, when the relative expression of WISP1 in either cell type was compared (Figure 4.6C), the WISP1 expression level in fibroblasts was approximately 17-fold higher than in ATII cells at baseline, and under hypoxic conditions WISP1 expression increased a further 5-fold.

Cooperation between TGF-β and HIF signalling has previously been observed in several tissue contexts, e.g. interactions between HIF1α and the TGF-β effector protein Smad3 linked to enhancement of erythropoietin expression via stabilisation of the interaction between promoter and enhancer DNA regions(225). This may have relevance to IPF, as both pathways are known to be active in IPF. However, figure 4.6 shows that this is unlikely to be occurring in the context of WISP1 expression in MRC-5 fibroblasts – there is no evidence of cooperation between the two signalling pathways in driving WISP1 expression in these cells. This may be because TGF-β mediated enhancement of WISP1 expression is inactive in these cells, or it may be that any TGF-β signalling is drowned out by the hypoxia stimulus, preventing any synergistic effect from being identified – i.e., that WISP-1 expression is already maximised due to the strong hypoxia stimulus. This is potentially supported by the synergistic upregulation of WISP1 expression by TGF-β and hypoxia signalling in ATII ER: KRAS V12 cells, as they express significantly lower levels of WISP1 at

baseline. Notably, however, it has been observed in primary fibroblasts derived from normal lung tissue that TGF-β treatment is inhibitory to IOX2-induced WISP1 expression. (Dr Elizabeth Davies, personal communication). Further experiments are required in order to investigate this finding, as it suggests that TGF-β signalling may inhibit HIF-induced WISP1 expression in lung fibroblasts.

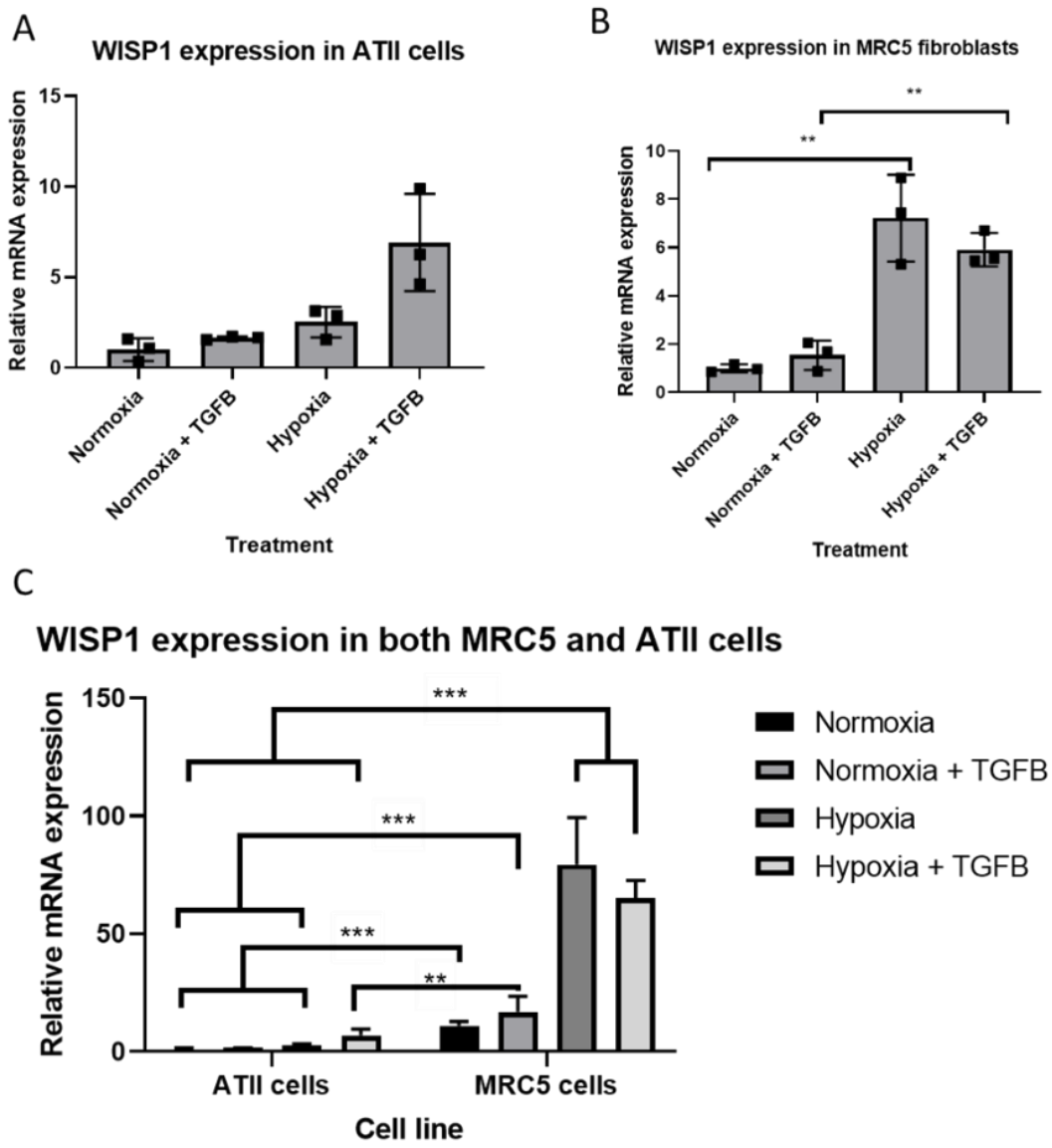


Figure 4.6: WISP1 expression in a cultured alveolar type II derived cell line (ATII ER: KRASV12) and MRC-5 cells. TGF-β treated cells were treated with 10ng/ml TGF-β for 24 hours. Hypoxia treated cells were grown in a hypoxia chamber (1% O_2 partial pressure) for 24 hours prior to harvest. Error bars show standard deviation. WISP1 expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC/GAPDH for ATII cells and UBC/A2 housekeeping gene expression C_t values. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. Significance stars represent results of one-way ANOVA, with *** denoting $p < 0.001$, **

denoting $p < 0.01$, and * denoting $p < 0.05$. A. WISP1 expression in an alveolar type II cell line, normalised to no TGF-β and normoxia in ATII cells. B. WISP-1 expression in MRC-5 lung fibroblasts normalised to no TGF-β and normoxia in MRC-5 cells. C. The same data as in A. and B. but normalised to no TGF-β and normoxia in ATII cells for both cell types.

4.8 WISP1 expression in primary lung fibroblasts is HIF1α-driven

Having identified WISP1 induction by hypoxia, the pathway by which this hypoxia signalling induces WISP1 was then investigated. HIF1α signalling is the primary means by which mammalian cells sense oxygen. HIF1α signalling is mediated via prolyl hydroxylase enzymes, which hydroxylate proline residues on various proteins, including HIF1α(226). Thus, a pilot experiment was performed using primary lung fibroblasts treated with siRNAs to knock down HIF1a before treatment with DMOG, a broad-spectrum prolyl hydroxylase inhibitor, to cause HIF stabilisation of activation of HIF signalling. Figure 4.7 shows the effect of inducing WISP1 expression with DMOG in the presence of siHIF1a or control siRNA. This result is only a single experiment, so is necessarily preliminary, but if it is replicable it demonstrates that HIF signalling is both necessary and sufficient to induce WISP1 gene expression in response to a PHD inhibitor. Whereas there was a strong induction of WISP1 expression in fibroblasts treated with DMOG and a control siRNA, this was almost completely abrogated by knockdown of the HIF1α gene, suggesting that the HIF1α signalling pathway induces WISP1 expression in healthy lung fibroblasts.

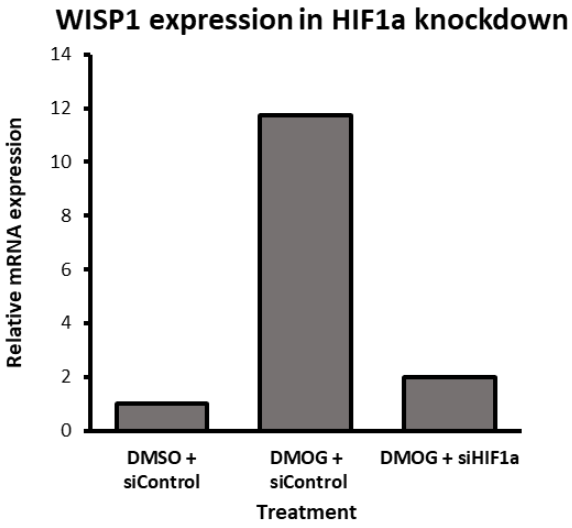


Figure 4.7: WISP1 expression as measured by qPCR in primary healthy lung fibroblasts treated with DMSO + control siRNA, DMOG and control siRNA, and DMOG and a HIF1α-targeting siRNA. DMOG concentration 1mM. WISP1 expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC and A2 housekeeping gene expression C_t values and DMSO and siControl C_t value. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. N = 1.

4.9 WISP1 expression in IPF fibroblasts is driven by IOX2, a selective PHD inhibitor

DMOG is a broad-spectrum prolyl hydroxylase inhibitor, and thus works to inhibit other PHD enzymes which are not directly related to sensing hypoxia – for example, there are several prolyl hydroxylases such as P4H which are important for collagen crosslinking(227). To mitigate this effect, and specifically target the PHD enzymes which hydroxylate HIF molecules, an alternative PHD inhibitor, IOX2, was used. This allows more specific targeting of hypoxia pathways while mitigating off-target effects.

In order to compare how well lung fibroblasts responded to IOX2 hypoxia analogue treatment, a dose-response experiment was performed using fibroblasts either from control or IPF lung tissue. Figure 4.8 shows a dose-response for IPF fibroblasts treated with 0, 10 and 50µM concentrations of IOX2, with significant differences in WISP1 expression relative to no IOX2 at both 50 and 250µM IOX concentrations. However, no such dose response was observed for normal, healthy lung fibroblasts, with their WISP1 expression trending downwards when treated with the highest IOX2 dose. This lack of a dose-response for the control fibroblasts suggests that IPF fibroblasts are

somehow potentiated to produce more WISP1 in response to HIF signalling, and that this potentiation is not present in normal lung fibroblasts.

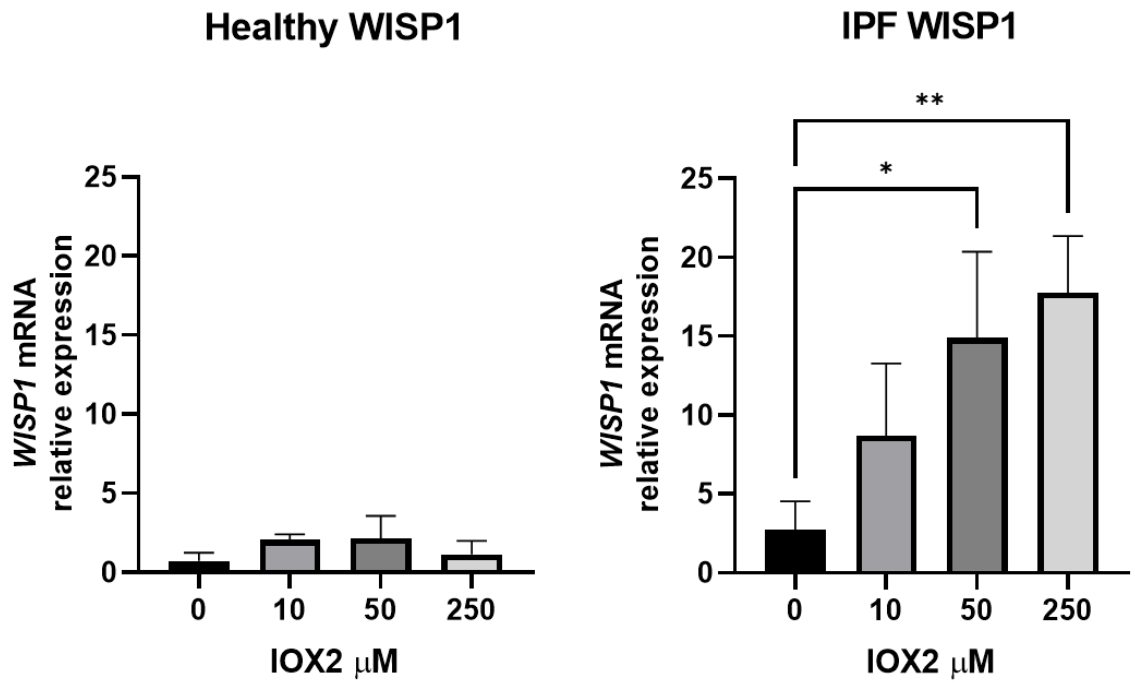


Figure 4.8: Dose response to IOX2 of WISP1 mRNA expression in normal and IPF cells. Cells were treated with a range of IOX2 doses. WISP1 expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC and A2 housekeeping gene expression C_t values and no IOX2 control C_t value for each cell line. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. Error bars are standard deviation. Significance stars represent results of a one-way ANOVA with Dunnett's multiple comparison test to compare means, with *** denoting $p < 0.001$, ** denoting $p < 0.01$, and * denoting $p < 0.05$. N = 3 for both fibroblast types.

4.10 WISP1 expression correlates with induction of collagen crosslinking enzymes in primary IPF fibroblasts

While WISP1 expression was induced by treatment of IPF fibroblasts with IOX2 to induce HIF signalling, the lack of effect of IOX2 on control fibroblasts was unexpected. Thus, the same cDNA preparations from Figure 4.8 were used to measure expression levels of the collagen crosslinking enzymes LOXL2 and PLOD2 in response to increasing doses of IOX2. These genes have been

previously associated with an increase in collagen fibril stiffness in IPF(65). shows these expression levels. Interestingly, while PLOD2 shows dose-response to IOX2 in both healthy and IPF lung fibroblasts, with both cell types showing an significant increase at the highest IOX2 dose, the same is not true for LOXL2, which showed a trend of induction in IPF fibroblasts even at 10µM IOX2, but considerably less sensitivity to IOX2 in healthy fibroblasts, in a manner similar to WISP1. However, the observed dose-response of LOXL2 expression to IOX2 is not significant in both cell types.

Although there are some differences in the dose-response of these genes to hypoxia analogue treatment, it is clear that in IPF fibroblasts, both WISP1 and PLOD2 are all upregulated by increased HIF signalling, while LOXL2 expression may be upregulated, but this was not significant. All three genes also co-express in IPF single-cell data, suggesting they may be coregulated by hypoxia signalling in IPF.

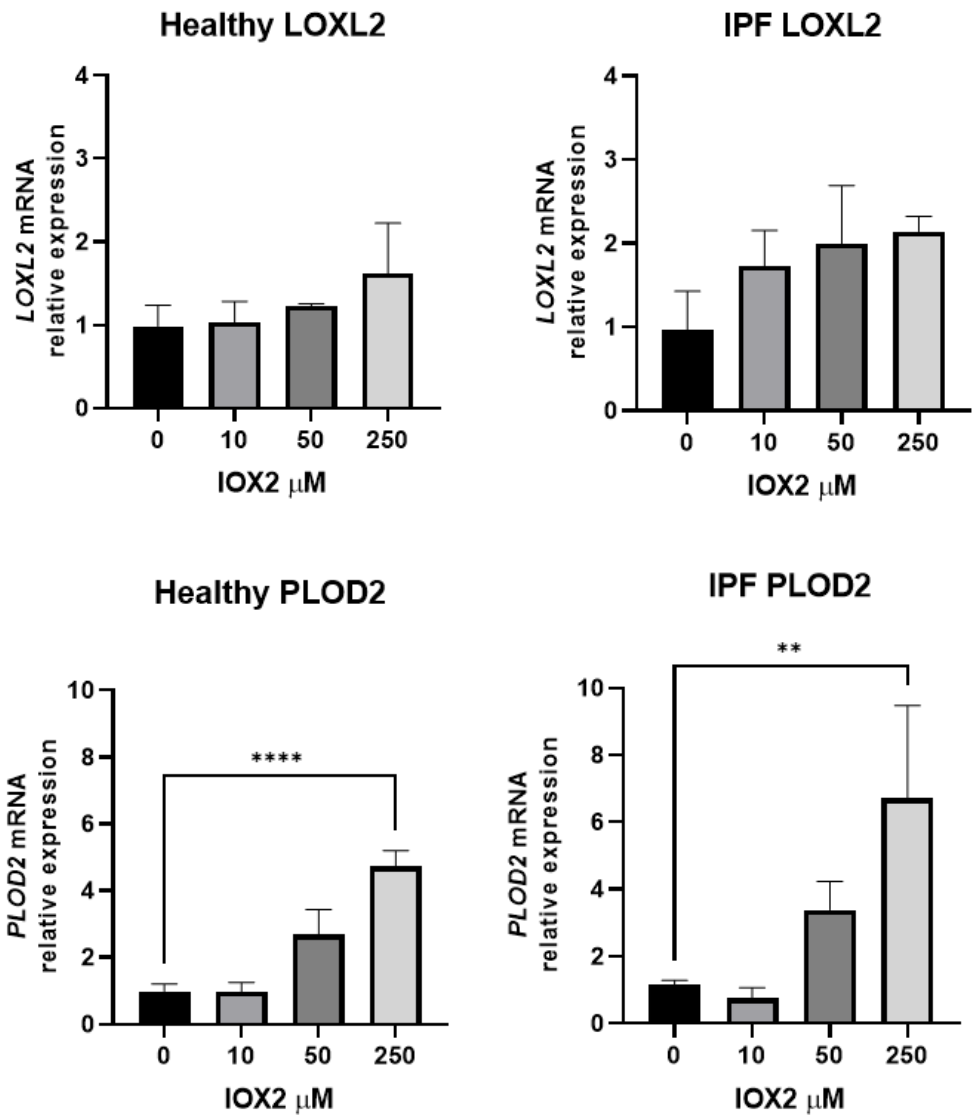


Figure 4.9: IOX2 dose-response of expression of the collagen crosslinking genes LOXL2 and PLOD2 in normal and IPF primary cultured lung fibroblasts. Gene expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC and A2 housekeeping gene expression C_t values and no IOX2 control C_t value in HL409 cells. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. N = 3 for both cell types. Significance stars represent results of a one-way ANOVA with Dunnett's multiple comparison test to compare means, with *** denoting $p < 0.001$, ** denoting $p < 0.01$, and * denoting $p < 0.05$. N = 3 for both fibroblast types.

4.11 WISP1 expression can be efficiently knocked down in both normal and IPF fibroblasts.

The data presented above, in shows WISP-1 expression is induced by IOX2 treatment in primary IPF fibroblasts. However, this induction was stronger in IPF fibroblasts compared with control fibroblasts (Figure 4.7).

In order to continue to investigate the upregulation of WISP1 expression in IPF fibroblasts, WISP1 expression was knocked down using siRNA. Figure 4.10 shows full-length WISP1 expression in two different primary lung fibroblast cell types, healthy (HL409) and IPF (K158) with higher WISP-1 expression being induced in the IPF fibroblasts. Figure 4.10 also shows very efficient knockdown of full-length WISP1 in the IPF fibroblasts, with the upregulation of WISP1 gene expression caused by the addition of IOX2 to the media almost completely abolished. These results confirm efficient targeting of WISP-1 by WISP-1 siRNA.

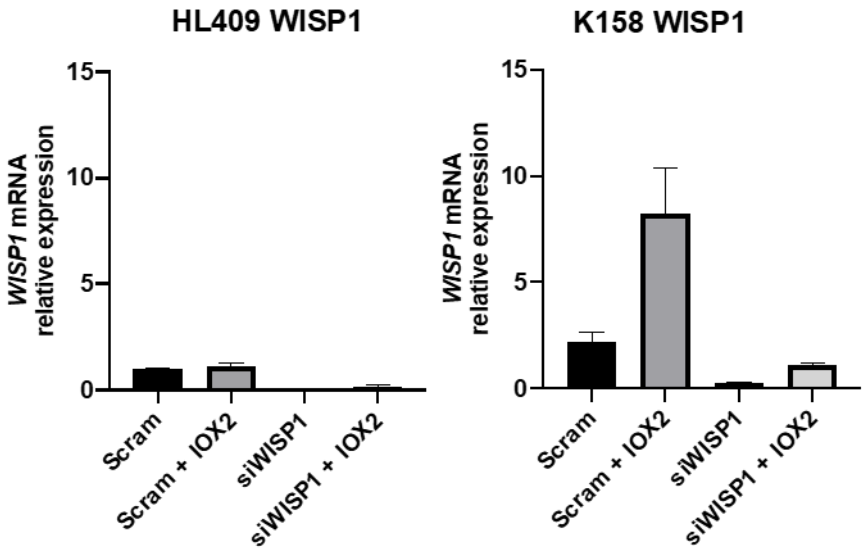


Figure 4.10: WISP1 mRNA expression in control (HL409) and IPF (K158) primary cultured fibroblasts treated with IOX2 and a WISP1 targeting siRNA smartpool. WISP1 expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC and A2 housekeeping gene expression C_t values and no IOX2 control C_t value in HL409 cells. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. Error bars are standard deviation. Cells were grown in media containing IOX2 (25µM) or DMSO. N = 2.

4.12 WISP-1 protein levels increase in IPF fibroblasts treated with IOX2.

Having identified WISP1 induction in response to HIF signalling and hypoxia at the mRNA, gene expression level, western blot analysis of WISP-1 protein levels was performed on cell lysates treated with IOX2 to identify if the WISP1 gene induction led to a commensurate increase in WISP-1 protein levels. Figure 4.11shows this analysis, with β-actin used as a loading control.

The findings shown here broadly mirror the findings identified for WISP1 mRNA expression levels. IPF fibroblasts have larger amounts of WISP-1 protein present at baseline, without HIF stimulation. WISP-1 protein induction by IOX2 treatment in the IPF fibroblasts was higher than in healthy fibroblasts, suggesting that the observed induction of WISP1 expression at the mRNA level also leads to an increase in WISP-1 protein production.

The results of the siRNA knockdown are less conclusive. Although the observed induction of WISP-1 in the IPF fibroblasts is almost entirely eliminated, the baseline level of WISP-1 protein does not change. This suggests that the duration of this experiment is not sufficiently long to allow the levels of WISP-1 protein to turnover and reflect the lower mRNA levels caused by the WISP-1 siRNA. A longer knockdown period would likely be required for more complete protein knockdown.

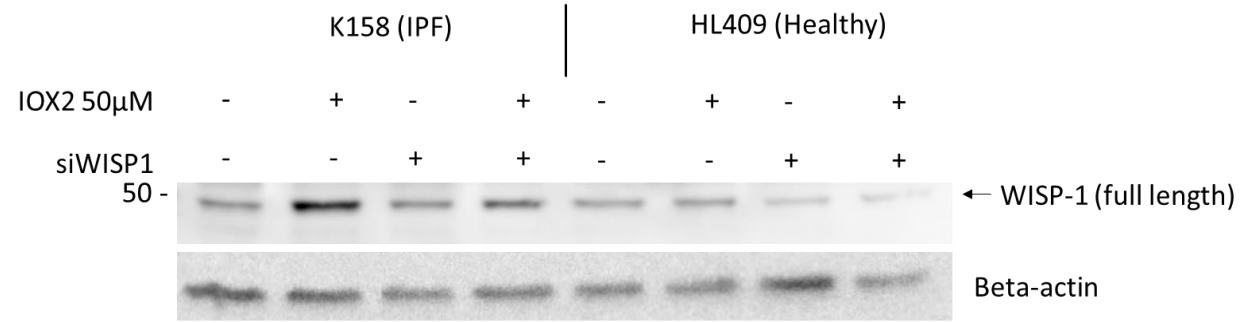


Figure 4.11: WISP-1 protein level in IOX2 treated primary normal (HL409) and IPF (K158) fibroblasts. Samples were treated with 25µM IOX2 for 24 hours.

4.13 Hypoxic conditions cause induction of WISP1 in IPF fibroblasts

Having established that WISP1 expression could be reliably induced by hypoxia analogues, and that this expression increase could be reduced by siRNA knockdown, cultured healthy and IPF lung fibroblasts were grown in a hypoxia chamber for 24 hours, to see if the same effect could be observed in response to hypoxia. Figure 4.12 shows the results of this treatment and knockdown. As seen above, the normal, healthy lung fibroblasts subjected to this treatment showed considerably lower expression of full-length WISP1 in these cells both at baseline and in response to hypoxia, suggesting that IPF cells are potentiated to WISP1 induction by hypoxia signalling, at least for the cells used in this experiment. As with Figure 4.10 above, WISP1-targeting siRNA treatment was sufficient to almost eliminate full-length WISP1 gene expression completely in these cells.

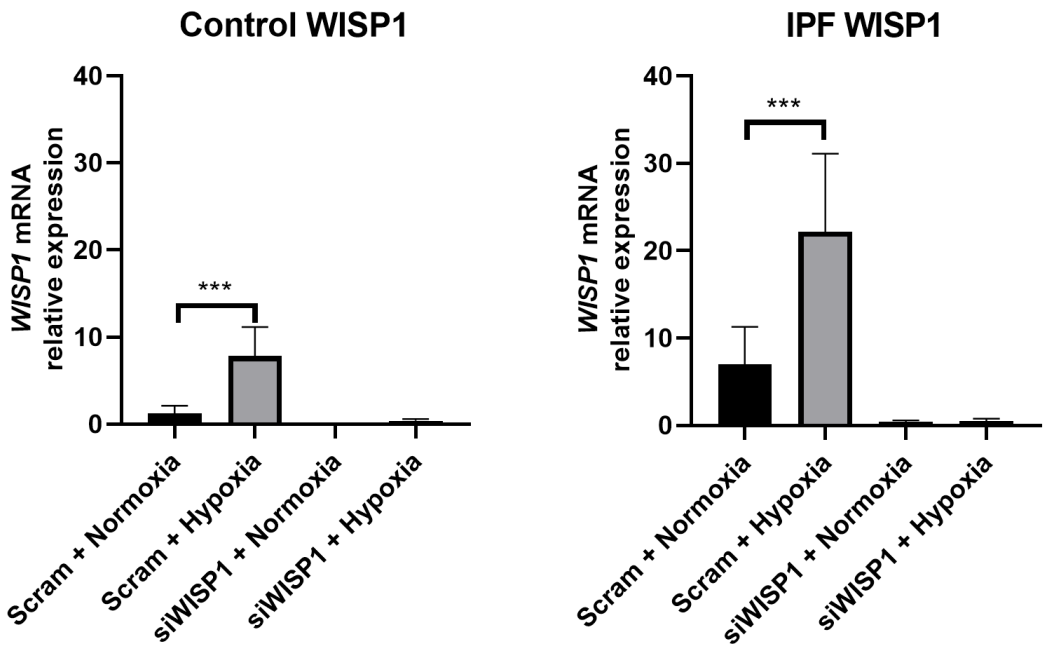


Figure 4.12: WISP1 mRNA expression in control (HL409) and IPF (K158) cells. Hypoxia treated cells were grown in a hypoxia chamber at a partial pressure of 1% O_2 for 48 hours before harvesting. Bars are the mean; error bars are standard deviation. WISP1 expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC and A2 housekeeping gene expression C_t values and scram + normoxia C_t value in control fibroblasts. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. N = 6 for both fibroblast types. (2 replicates of 3 cell lines). Significance stars represent results of a t-test, with *** denoting $p < 0.001$, ** denoting $p < 0.01$, and * denoting $p < 0.05$. N = 6 for both fibroblast types.

Having corroborated the qPCR data at the protein level for WISP1 gene expression induction by IOX2-induced HIF signalling, the same was done for hypoxia treated fibroblasts. Figure 4.13 shows the same pattern of WISP-1 protein levels seen in Figure 4.11, as well as the same effect on siWISP1 treated cells; thus siWISP1 treatment prevented the increase in WISP-1 protein seen with hypoxia treatment, but there was a relatively unchanged baseline expression level.

The western blots shown here only show full-length WISP-1 protein. The splice variants identified in Figure 4.16 should all be detectable by the primary antibody used for this blot, with the exception of variant 3, as it contains a truncated C-terminus which lacks the C-terminal epitope used to generate the antibody. However, only bands corresponding to the full-length protein were visible, perhaps reflecting the lower baseline level of the WISP-1 splice variants compared to the full-length protein in these cells.

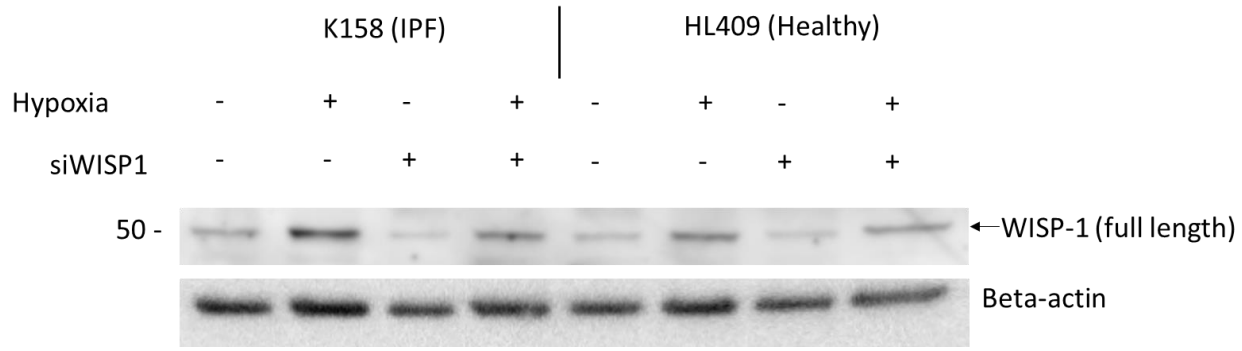


Figure 4.13: WISP-1 protein level in hypoxia treated primary normal (HL409) and IPF (K158) fibroblasts. Samples were grown in hypoxic conditions (1% O_2 partial pressure) for 24 hours prior to harvesting.

4.14 PPARG, a previously identified WISP-1-responsive gene, is upregulated in response to WISP1 knockdown

Having established that WISP1 can be efficiently knocked down in IPF fibroblasts, the functional consequences of this knockdown were explored. A previous study looking at how WISP1 regulates bone and fat formation in perivascular stem cells identified that WISP1 knockdown leads to an increase in PPARG expression(155). PPARG codes for Peroxisome Proliferator Activated Receptor Gamma (PPARγ). PPARγ regulates fatty acid storage in adipocytes, and is also characteristically expressed in rodent lipofibroblasts(54,153,228). Another gene which was identified as being downregulated in this study in response to WISP1 knockdown was Runt-related transcription factor 2 (RUNX2), a transcription factor associated with bone development and osteoblast differentiation(229). This paper found that WISP1 is associated with upregulation of a bone-type phenotype in perivascular stem cells, coupled with downregulation of an adipogenic phenotype(155).

Figure 4.14 shows how WISP1 knockdown and hypoxia affected the expression of these two genes. Notably, PPARG expression was significantly increased in both control and IPF fibroblasts in which WISP1 had been knocked down, suggesting that suppression of adipogenic signalling may be a common feature of WISP-1 signalling. However, RUNX2 expression was not affected by WISP1 knockdown in either healthy or IPF fibroblasts, although it was significantly decreased under hypoxic conditions in IPF fibroblasts. In summary, expression of PPARG was suppressed by WISP1 irrespective of the O2 tension, whereas RUNX2 expression was not affected by WISP1 modulation, but was decreased when the O2 tension was lowered. These data suggest that PPARG and RUNX2 are differently regulated in lung fibroblasts.

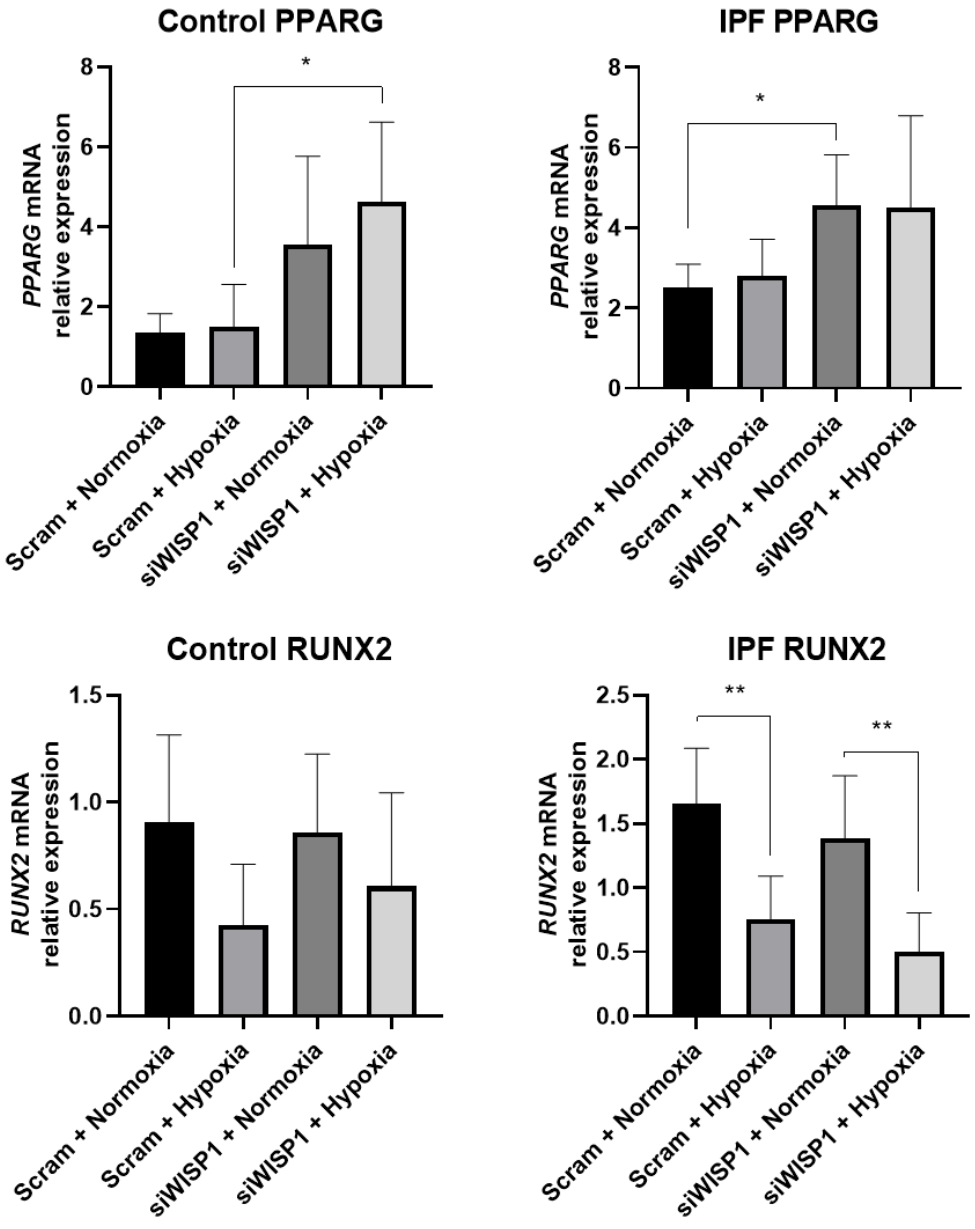


Figure 4.14: Expression of PPARG and RUNX2 genes in control and IPF fibroblasts in normoxic and hypoxic (partial pressure O_2 = 1%) conditions with WISP1 gene knockdown. Gene expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC and A2 housekeeping gene expression C_t values and scram + normoxia C_t value in control fibroblasts. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. Significance stars represent results of one-way ANOVA, with *** denoting $p < 0.001$, ** denoting $p < 0.01$, and * denoting $p < 0.05$. Error bars are standard deviation. N = 6 for both fibroblast types (2 replicates of 3 cell lines).

4.15 Expression of WISP1 splice variants is induced by hypoxia in cultured IPF fibroblasts

All WISP1 results presented above related to the expression of full-length WISP1 which contains all five possible WISP1 domains. However, there are three other alternatively splice variants which are capable of producing viable protein, each lacking different domains and thus possibly affecting WISP-1 functionality(156,157,230). In order to determine if these variants are differentially regulated by hypoxia signalling, their expression levels were assessed by qPCR in control and IPF lung fibroblasts. Their expression in siRNA WISP1 knockdown samples was also investigated. AS shown in Figure 1.14, the siRNA Smartpool used to knock down WISP1 expression targeted multiple areas of the WISP1 gene, including areas on all four WISP1 variants, so the siRNA methodology used was expected to be capable of knocking down expression of all splice variants.

Figure 4.15: Targeting of WISP1 splice variants with Smartpool siRNA. Highlighted areas show areas where siRNA binds. V1 is full-length WISP1.

Figure 4.16 shows the effect of hypoxia on expression of the different splice variants in normal and IPF lung fibroblasts grown under hypoxic conditions. Notably, all three truncated WISP1 variants exhibited some induction with hypoxia, with variants 2 and 4 showing particularly strong hypoxia-induced expression. The siRNA knockdown was also very effective for all variants, suggesting that the multiple target loci of the different constituent Smartpool RNA molecules (Figure 4.15) allowed efficient targeting of all WISP1 splice variants. Interestingly, there was comparatively lower baseline expression, and lower induction, of all three splice variants in healthy fibroblasts compared to IPF fibroblasts, suggesting that the potentiation for induction of WISP1 expression in IPF fibroblasts seen with the full-length transcript is also present in IPF fibroblasts.

Notably, however, the baseline expression level of all three truncated splice variants was lower than for the full-length gene, suggesting that full-length WISP1 is the dominant form of the protein in IPF fibroblasts.

Although it was possible to detect full-length WISP-1 protein by Western blotting, the presence of the splice variant proteins was not confirmed, even though variants 2 and 4 should be detectable by the primary antibody used for western blotting. However, variant 3 which lacks the C-terminal epitope recognised by the antibody, would not be detected. Failure to detect the splice variants may reflect the lower baseline level of the WISP-1 splice variant mRNAs compared to the full-length protein in these cells. However, further work is required to confirm these findings.

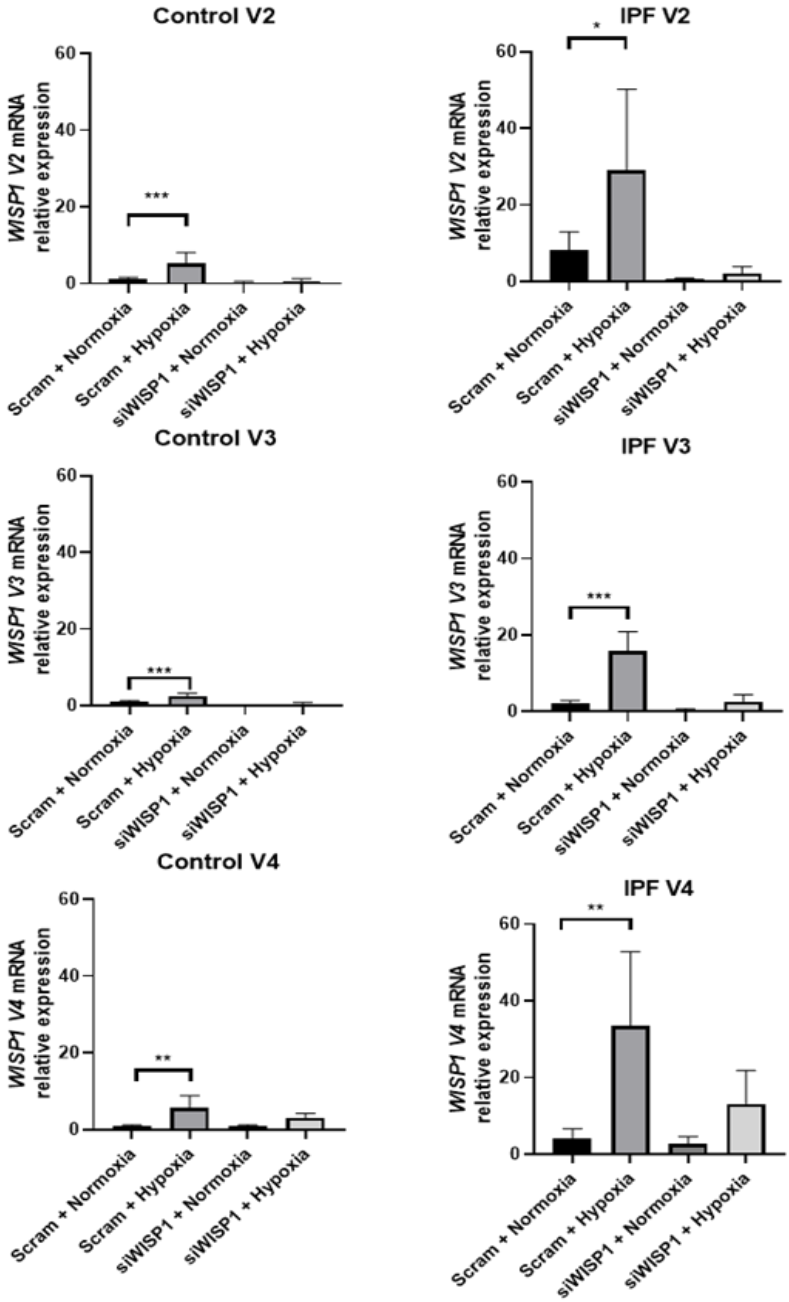


Figure 4.16: WISP1 splice variant expression in Control and IPF primary fibroblasts. Splice variant-specific qPCR primers were used to measure the expression level of different splice variant mRNA. Bar height represents mean of 6 replicates. WISP1 splice variants expression was quantified using real time quantitative PCR (RTqPCR). $\Delta\Delta C_t$ values were calculated by normalisation to UBC and A2 housekeeping gene expression C_t values and scram + normoxia C_t value in control fibroblasts. Relative mRNA expression values are $2^{-\Delta\Delta Ct}$. Error bars are standard deviation. Significance stars represent results of one-way ANOVA, with *** denoting $p < 0.001$, ** denoting $p < 0.01$, and * denoting $p < 0.05$.

4.16 Conclusions

The research presented in this chapter demonstrates that WISP1 gene expression can be induced by hypoxia, and by stimulating the HIF signalling pathway using two different prolyl hydroxylase enzyme inhibitors. Knockdown of HIF1α in cultured primary lung fibroblasts almost completely removed WISP1 gene expression, suggesting that WISP-1 is induced by HIF1α signalling. The observed induction of full-length WISP1 was also observed at the protein level in lysates from cells treated with IOX2 and grown in hypoxic conditions, suggesting that the induction of WISP1 gene expression does translate to a commensurate increase in protein production. The gene expression levels of WISP1 splice variants were also measured and were also increased by culturing cells under hypoxic conditions.

Induction of WISP1 gene expression was higher in IPF versus healthy lung fibroblasts. Therefore, these data suggest that IPF fibroblasts may be potentiated for increased WISP-1 production in response to hypoxia. The baseline level of WISP1 expression was also higher in cultured primary IPF lung fibroblasts than in healthy fibroblasts. This may suggest that basal HIF activity is higher in IPF fibroblasts, even in normoxic conditions.

This chapter also demonstrates the feasibility of performing WISP1 knockdown studies. In all conditions identified, the increase in WISP1 gene expression in cultured lung fibroblasts was substantially reduced with the addition of siRNA, and WISP1 gene expression was very substantially reduced in general. This also extends to efficient knockdown of WISP1 splice variants. However, protein levels were less affected by gene knockdown, suggesting that the baseline level of WISP-1 requires longer than 48 hours to turnover and be removed. The increase in WISP-1 protein levels in response to IOX2/hypoxia was abrogated, however.

Although induction of WISP1 by HIF activation in control fibroblasts was low, IOX2 treatment did significantly induce expression of PLOD2, which is known to be HIF-regulated, to comparable levels in control and IPF fibroblasts. However, for WISP1 there was a difference between the dose responses and sensitivities of control and IPF fibroblasts, suggesting additional mechanisms by which it may be controlled. LOXL2 expression was not significantly changed by IOX2 treatment in either cell type, but it did show an increasing trend in primary IPF fibroblasts. Nonetheless, these data confirm the observed correlation between WISP1 gene expression and collagen crosslinking gene expression shown in chapter 3, at least for WISP1 and PLOD2.

Knockdown of WISP1 led to an increase in PPARG expression. This may indicate WISP-1 functions to suppress lipogenic phenotypes in fibroblasts. However, expression of RUNX2, an osteogenic

factor, was unchanged with WISP1 knockdown and, interestingly was suppressed by hypoxia. Assuming WISP1 and HIF signalling play a role in IPF pathogenesis, these data suggest that other factors are required to maintain or increase RUNX2 expression, or that RUNX2 is not important for this mechanism.

Altogether, the research presented in this chapter presents a clear case that WISP1 is induced by hypoxia, and specifically by HIF signalling, likely through HIF1α. This effect was more pronounced and consistent in IPF fibroblasts, which also had a higher baseline expression level of WISP1. Other profibrotic processes, including TGF-β signalling and Wnt signalling had little effect on WISP1 gene expression, despite mRNA belonging to the noncanonical Wnt signalling ligand WNT5A also localising to fibroblast foci. Wnt-inducible signalling protein 1 is therefore perhaps a misnomer, at least in the context of the role of WISP-1 in IPF. WISP1 gene expression can also be efficiently knocked down in both healthy and IPF fibroblasts, even when those fibroblasts are grown in hypoxic conditions or treated with IOX2, a prolyl hydroxylase inhibitor which mimics the effect of HIF signalling. WISP1 knockdown also has functional consequences in both healthy and IPF fibroblasts, leading to increased PPARG expression.

4.17 Chapter Discussion

Much work has focussed on identifying the signalling pathways which are important for driving fibrosis in IPF. TGFβ signalling is the prototypical profibrotic signalling pathway and has been identified as being important for IPF pathogenesis. TGFβ signalling in IPF has been linked to epithelial to mesenchymal transition of lung epithelial cells, leading to them developing a more mesenchymal phenotype, recruitment of circulating fibrocytes to the IPF lung, and fibroblast proliferation and differentiation leading to the myofibroblasts present in fibroblastic foci. TGFβ signalling is therefore well-integrated into the current model of IPF pathogenesis, where tissue microinjuries promote chronic activation of wound-healing processes, leading to recruitment, activation and proliferation of ECM-depositing myofibroblasts, leading to the systemic changes to the interstitium which are characteristic of the disease.

Other pathways known to be important in IPF are Wnt signalling and hypoxia. Wnt signalling is a pathway primarily investigated for its role in development, but it is also activated in wound healing, and has been implicated in IPF pathogenesis.

Hypoxia signalling via the HIF proteins has been heavily investigated as contributing to multiple different diseases, notably cancer. Several studies have identified it as being important in IPF, with HIF1α being overexpressed in microarray studies of IPF tissue(94), stimulating epithelial to

mesenchymal transition(110), and promoting fibroblast to myofibroblast differentiation in IPF(95). HIF1α signalling has also been implicated in liver fibrosis, with a mouse model of liver fibrosis showing reduced fibrogenesis in HIF1α-deficient mice versus controls.(231) Recent research has also identified hypoxia as upregulating collagen crosslinking enzymes in IPF, suggesting a mechanism by which it may contribute to fibrosis.

The finding that WISP-1 is HIF responsive has important implications for its functional role in fibrosis. It is likely that WISP-1 production increases with worsening fibrosis, as this will lead to higher tissue hypoxia as gas exchange is further impaired. Instead of WISP-1 being associated with wound-healing processes, it is upregulated by pathogenic hypoxia.

WISP1 expression is higher at baseline in IPF fibroblasts compared to healthy fibroblasts. WISP1 expression is also more sensitive to IOX2 concentration in a dose-response This may suggest that the baseline level of HIF signalling is higher in IPF fibroblasts even in normoxic conditions. Although hypoxia is the primary means of inducing HIF signalling, other processes can regulate HIF turnover as well. Notably, factor inhibiting hypoxia (FIH) regulates HIF activity in response to oxidative stress(232). Oxidative stress has been implicated in IPF pathogenesis(233). Increased FIH activity in response to oxidative stress could lead to increased HIF activation in IPF, leading to increased baseline WISP1 expression.

The lack of response of WISP1 gene expression to Wnt signalling, despite most previous research on WISP-1 focussing on its induction by Wnt signalling, is of interest. This suggests that regulation of WISP-1 expression may be context dependent, and that regulation by Wnt signalling may be less relevant to its function in IPF. It could also suggest that Wnt signalling alone is insufficient to induce WISP1 expression. The RNAscope *in-situ* hybridisation demonstrated that WNT5A is expressed in IPF tissue, and that fibroblastic foci are enriched for WNT5A expression as well as being the sites of WISP1 expression in IPF. It is likely Wnt signalling is active in fibroblastic foci. It may be that Wnt signalling activation is insufficient by itself to induce WISP1 expression in IPF fibroblasts, but in combination with other signalling pathways, such as HIF signalling, it upregulated WISP1 expression. This could be an alternative explanation for the increased baseline expression of WISP1 in IPF fibroblasts, or the potentiation of IPF fibroblasts to HIF induction of WISP1 expression. Conversely, it suggests that the upregulation of WISP1 in other diseases may be at least partially due to an increase in hypoxia signalling, particularly in reference to WISP1 being an oncogene responsible for increased cellular proliferation in cancer – hypoxia being an important process in the development of many cancers as tumours are often oxygen-poor environments(131).

It is evident that the IPF fibroblasts tested here showed greater hypoxia/HIF induced WISP1 expression versus the healthy control fibroblasts. They also had somewhat greater baseline WISP1 expression levels, as well as higher base WISP-1 protein levels. This suggests that IPF fibroblasts are potentiated to express more WISP1 compared to healthy lung fibroblasts. It is unclear why this is the case – they may be more susceptible to HIF signalling in general, or there may be fewer downregulators of WISP1 expression in IPF fibroblasts which dampen its upregulation due to hypoxia signalling. In either case, it suggests that there is some pathological component to fibroblasts in IPF which makes them better able to produce WISP1.

The almost complete elimination of WISP1 gene expression in IOX2-treated fibroblasts in which HIF1α was knocked down specifically points to HIF1α as being responsible for HIF-induced WISP1 expression. HIF1α and HIF2α both upregulate overlapping genesets, but their functions are distinct, with HIF1α being more responsible for acute hypoxia sensing, while HIF2α upregulates genes associated with chronic hypoxia. Further experiments to see if this dependency is different in IPF fibroblasts exposed to chronic hypoxia would be of interest, allowing investigation of the specific hypoxia-regulated signalling pathway or pathways which induce WISP1 expression. This experiment should also be repeated with a larger number of primary fibroblast lines from both healthy and IPF tissue.

The upregulation of different WISP1 splice variant genes in this data is of interest, as all three alternatively spliced WISP1 variants lack certain domains which may mediate protein interactions, and thus the functionality of the WISP-1 protein. There was significant upregulation of the splice variants something that has been mooted as being a potential mechanism for WISP1 contributing to pathogenesis(143). However, the WISP-1 splice variants, even those which should have been detectable using the primary antibody used, were undetectable on a western blot, suggesting that the protein levels of WISP-1 splice variants is substantially lower than that of the full-length protein. This could be explained by preferential secretion of these variants – this could be determined by performing an anti-WISP1 western blot on media taken from cells with active HIF signalling.

Functional analysis of WISP1 knockdown was limited to quantifying expression of PPARG and RUNX2. Both of these genes have been shown to be modulated by WISP1 knockdown previously, and both link to previously established roles of WISP1 signalling in other tissue types – WISP1 is a known adipokine, as well as being important for bone development and osteogenesis(150,155,234). RUNX2 expression was not affected by WISP1 knockdown, however PPARG expression was significantly increased with WISP1 knockdown. PPARγ is a marker gene for

lipofibroblasts, so its regulation by WISP-1 may be evidence of WISP-1 suppressing lipofibroblast-like phenotypes in IPF fibroblasts(54,153). Rather than quantifying multiple potential WISP-1 regulated genes, however, it was decided to take this work forward separate from this thesis by performing RNAseq analysis on WISP1 knockdown samples, to quantify transcriptome-wide phenotypic changes brought about by WISP1 knockdown. WISP1 knockdown has been shown to have diverse effects in different cell types – it is associated with upregulation of EMT markers in melanoma, for example, but is associated with upregulation of bone morphogenic protein 3 (Bmp3) in mesenchymal stem cells(235,235).

Although the data presented in this chapter requires additional work to cement the role of WISP1 in affecting gene expression in IPF, it provides strong evidence that HIF activation is the cause of WISP-1 upregulation in IPF. Given the known role of HIF pathways in contributing to IPF pathogenesis, this meshes well with the idea that WISP-1 is a profibrotic gene and explains the observed correlation of WISP-1 expression with several profibrotic genes shown in chapter 3.

Future work for investigating functional roles of WISP-1 in IPF should centre on growing IPF fibroblasts under hypoxic conditions or with small molecule hypoxia analogues, and identifying any effects which WISP1 gene knockdown has on the cells' phenotype, either through functional assays or quantification of gene expression (i.e., using RNAseq).

5 Investigating functional roles of WISP-1 protein interaction partners

5.1 Chapter introduction and Aims

Having identified the localisation, context and likely drivers of WISP-1 expression in IPF, it remained to identify potential functional roles of WISP-1 in this disease. As WISP-1 is a member of the CCN protein family, known to mediate much of their function via protein interactions, and as previous functional roles of WISP-1 have been tied into its protein interaction partners, it was hypothesised that identifying interaction partners of WISP-1 will give insight into its functional roles. The first aim of this chapter was to generate GFP-tagged WISP-1 expression constructs and confirm their expression in HEK-293T and MRC-5 cells. The second aim was to use WISP-1 co-immunoprecipitation using a GFP-trap kit, and mass spectrometry of the resulting captured proteins to identify WISP-1 interaction partners, in an affinity purification-mass spectrometry workflow. Finally, this chapter looks at verification of these interaction partners and exploration of their functional roles in the context of IPF fibroblasts.

5.2 Generation of GFP tagged expression constructs

In order to investigate interaction partners of WISP-1, tagged, overexpression constructs of WISP-1 were generated. Green fluorescent protein (GFP) was used as the tag, both because its fluorescence can be readily identified under a fluorescent microscope, allowing visual confirmation of transfection success and efficiency, and also because it allows the use of high-affinity anti-GFP immunoprecipitation kits. This approach bypasses potential issues of low affinity and off-target effects previously described with commercial anti-WISP-1 antibodies(143).

An overexpression construct using a constitutively active CMV promoter was chosen to maximise the chances of WISP-1 protein-protein interactions being detected, as the physiological expression level of WISP1 is quite low. This comes with some risk of identifying non-specific/non-physiologically relevant interactors, however. All splice variants of WISP1 were incorporated into separate vectors, labelled variants 1-4

The polymerase chain reaction (PCR) based cloning strategy to produce these constructs is detailed in Figure 5.1. Briefly, DNA sequences overlapping the 5’ end of the GFP gene and the 3’ end of the WISP1 gene were added using primers with the appropriate sequences, as detailed in the Methods section. Primers were also used to insert Nhe1 and Kpn1 restriction sits on the 5’

and 3' ends of the fusion construct respectively. A fusion PCR reaction was then performed using the two overlapping WISP1 and GFP DNA fragments to produce a tagged expression construct, and this was then amplified, digested using Nhe1 and Kpn1, and ligated into the pcDNA 3.1(+) mammalian expression vector (Figure 5.2).

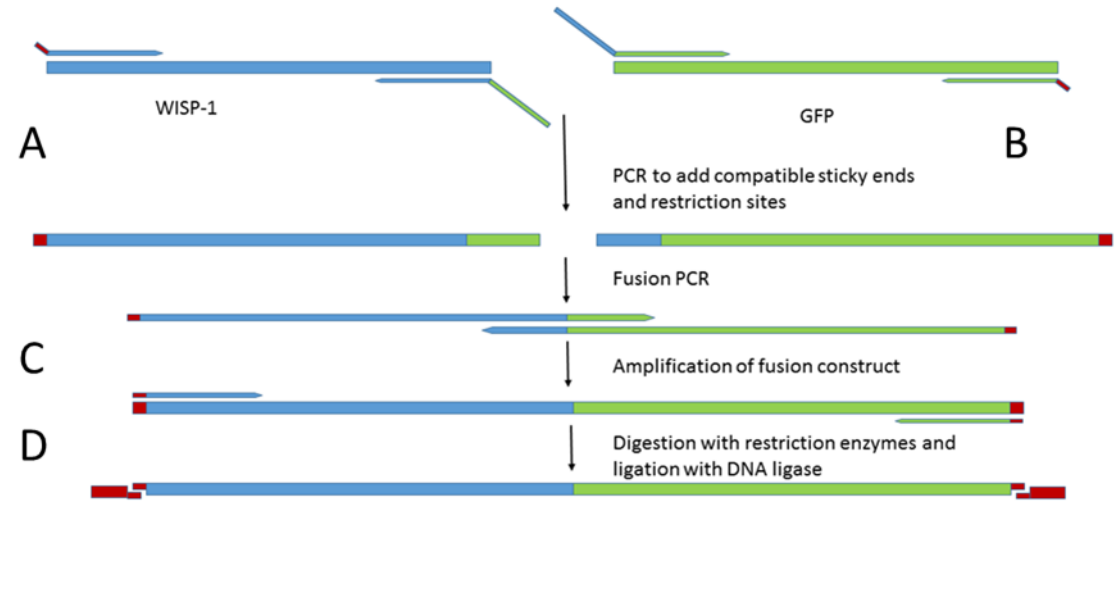


Figure 5.1: Schematic of how GFP-tagged expression constructs were generated and cloned. PCR was used to amplify WISP1 and GFP gene sequences, using primers designed to overlap with the end of the other gene sequence. The two products were then combined and PCR used to produce a fusion product containing WISP1 and GFP. This is followed by restriction digestion and ligation into pcDNA3.1(+). Letters refer to the steps described in Figure 5.3.

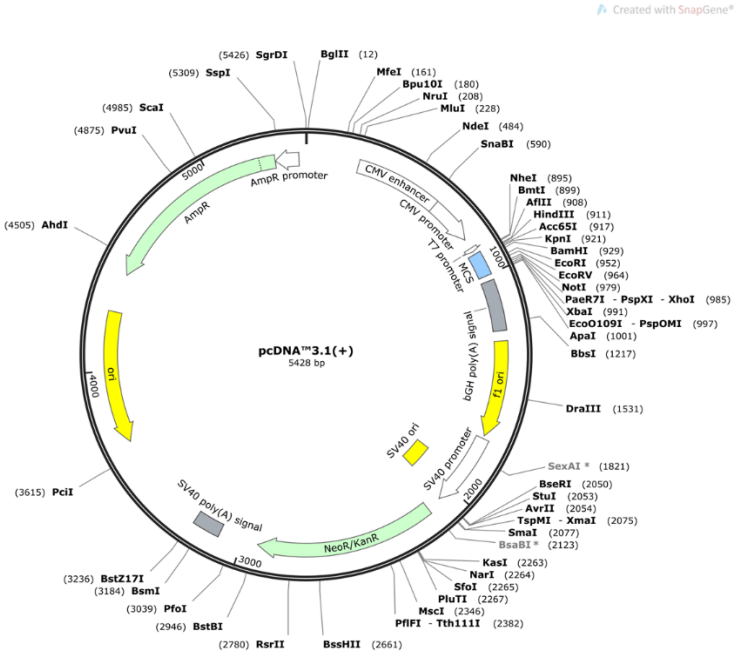


Figure 5.2: Map of pcDNA3.1(+) expression vector, showing location of restriction endonuclease cleavage sequences, antibiotic resistance genes.

5.3 Verification of expression constructs

The PCR products were verified agarose gel electrophoresis, and PCR fragments purified using a DNA gel purification kit (Qiagen). These purified fragments were then used for subsequent PCR reactions or restriction enzyme digestion/ligation. Complete plasmid was amplified in *E. coli* and purified using Qiagen's miniprep kits, then verified using restriction enzyme digestion of WISP-1 followed by agarose gel electrophoresis to confirm the insert size. Expected insert sizes were 1.8kb for full-length WISP1-GFP, 1.6 kb for variant 2-GFP, 1.2kb for variant 3-GFP and 1.1kb for variant 4-GFP. The plasmid was then sequenced to confirm the insert sequence was correct.

Figure 5.3 shows agarose gels containing PCR products and restriction digested plasmids used to generate and confirm these expression constructs.

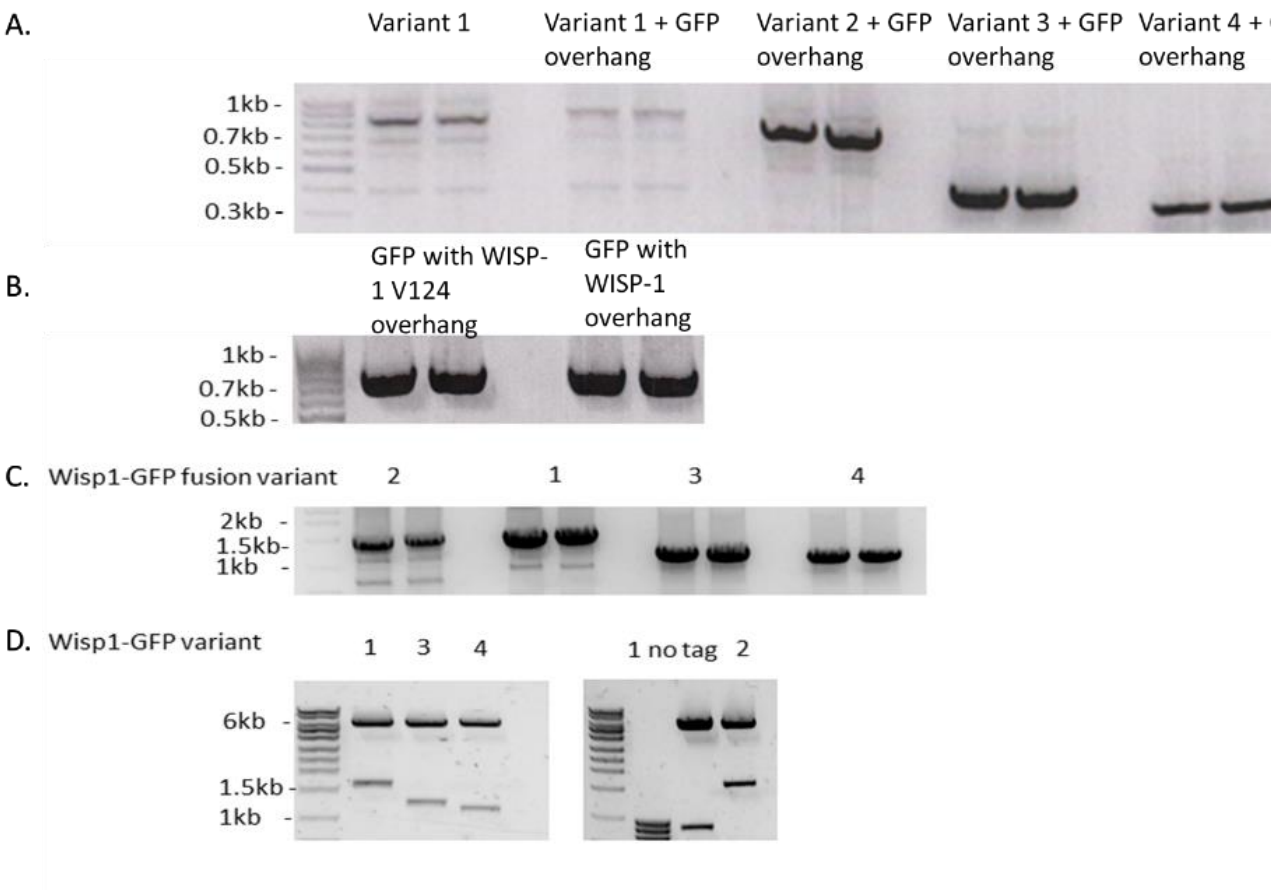


Figure 5.3: Generation of GFP-tagged WISP-1 constructs. A. DNA electrophoresis gel showing PCR products for each WISP-1 variant. Overhang refers to the use of primers that add a sequence overlapping the 5' end of the GFP gene, allowing fusion PCR with the GFP gene. B. DNA gel showing PCR products for GFP amplified using primers with overlap with the 3' ends of WISP-1 variant 1,2 and 4 (right lane), and variant 3 (left lane). C. DNA gel of products from fusion PCR of WISP-1 variants fused with the GFP gene. D. Diagnostic digests of plasmid maxipreps after restriction ligation and transformation using Kpn1 and Nhe1 restriction enzymes showing fragment sizes corresponding to GFP-tagged WISP-1 constructs. Constructs were also sequenced to verify their sequence. Confirmation of expression of WISP-1 tagged expression constructs shows strong fluorescence in HeLa and HEK293T cells

5.4 Tagged WISP-1 expression constructs are readily expressed at high efficiency in HeLa cells

In order to ensure that constructs were efficiently expressed, they were initially transfected into HeLa cells, both for their ease of transfection as well as strong adherence to coverslips, allowing cells to be formalin fixed and imaged using the confocal fluorescence microscope. GFP fluorescence was detected at an emission wavelength of 510nm. Variants 1, 3 and 4 were imaged using the fluorescence microscope, with variant 2 expression being subsequently verified in HEK293T cells. Figure 5.4 shows confocal microscopy images of HeLa cells transfected with the different GFP-tagged splice variants of WISP-1, as well as a GFP containing positive control vector. Notably, the subcellular localisation of the different splice variants shows all three variants shown to have some nuclear localisation in these cells. While this may be an artefact of the abnormally high WISP-1 levels from the plasmids, or the presence of the tag, it suggests a mechanism by which WISP-1 could directly influence gene transcription. Variant 3 also shows evidence of aggregation, with point-like areas of high fluorescence.

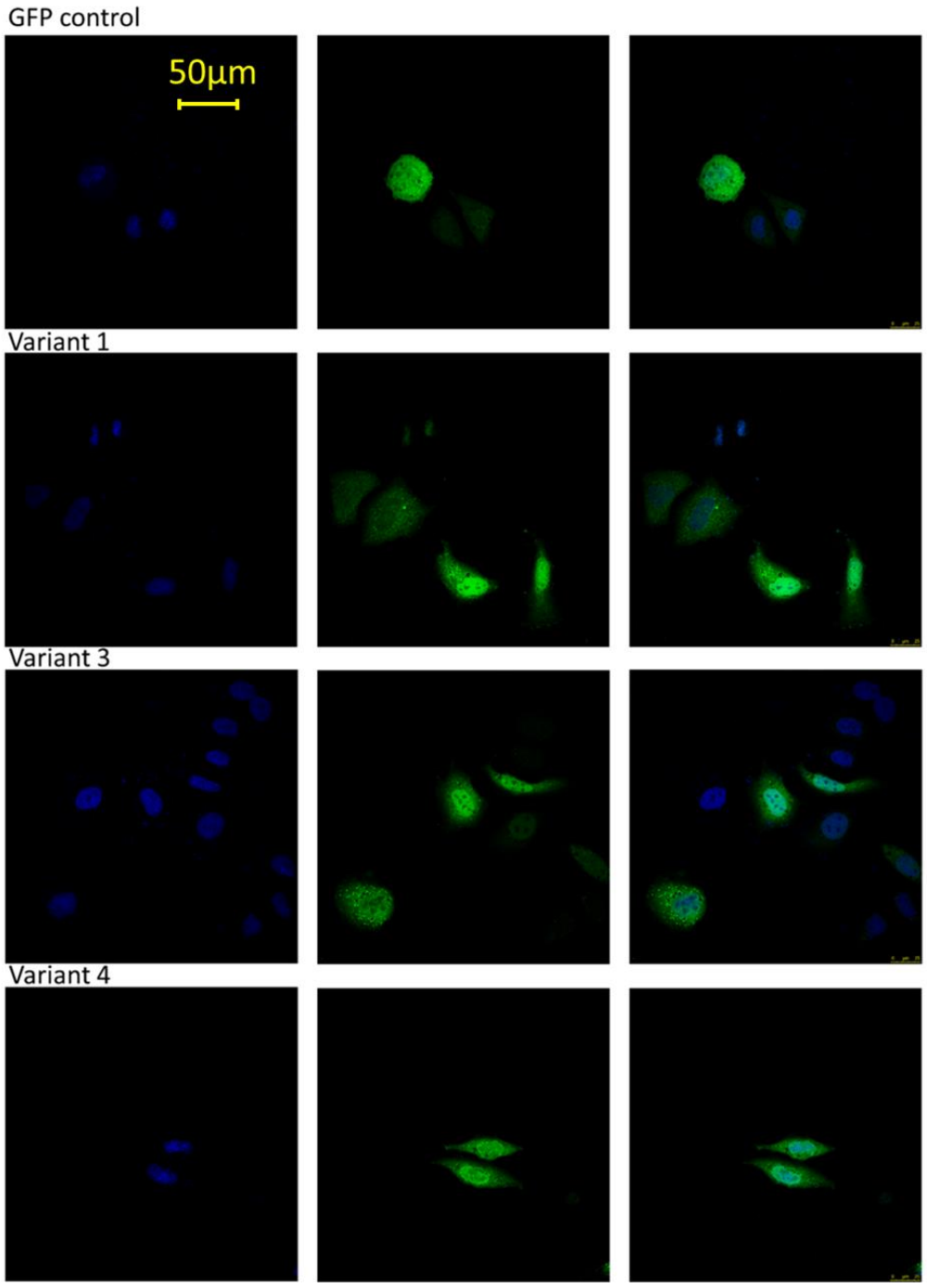


Figure 5.4: Confocal microscopy of a single optical section of HeLa cells transfected with constructs expressing GFP-tagged WISP-1 variants. Labels refer to the variants being expressed in cells, with the top set of images showing GFP alone. Blue fluorescence is DAPI (461nm emission), green fluorescence is GFP (509nm emission). Right-most images show DAPI and GFP images combined.

5.5 Fluorescence microscopy of HEK293T cells transfected with GFP-tagged WISP-1 shows expression of all 4 variants and GFP itself.

Although HeLa cells were readily transfected, and more amenable to confocal microscopy due to being better able to adhere to cover slips, transfection and initial co-immunoprecipitation experiments were performed in HEK293T cells. These are exceptionally amenable to transfection at high efficiency with minimal cell death, as well as being fast-growing and easy to culture. All four WISP-1 variants were able to be transfected in HEK293T cells at approaching 100% efficiency. Figure 5.5 shows fluorescence microscopy images of HEK293T cells expressing GFP-tagged versions of all WISP-1 variants, as well as GFP by itself. Cells were transfected using TransIT-X2 reagent at a 3:1 reagent: DNA ratio. Transfection of all vectors was very efficient and resulted in readily detectable fluorescence. There was little difference in fluorescence intensity between cells transfected with the different GFP-tagged WISP1 expression constructs. However, the same vector with GFP alone showed substantially stronger fluorescence intensity than the WISP1-GFP containing vectors, suggesting WISP-1-GFP is degraded faster than GFP alone, exported or translated less efficiently in these cells.

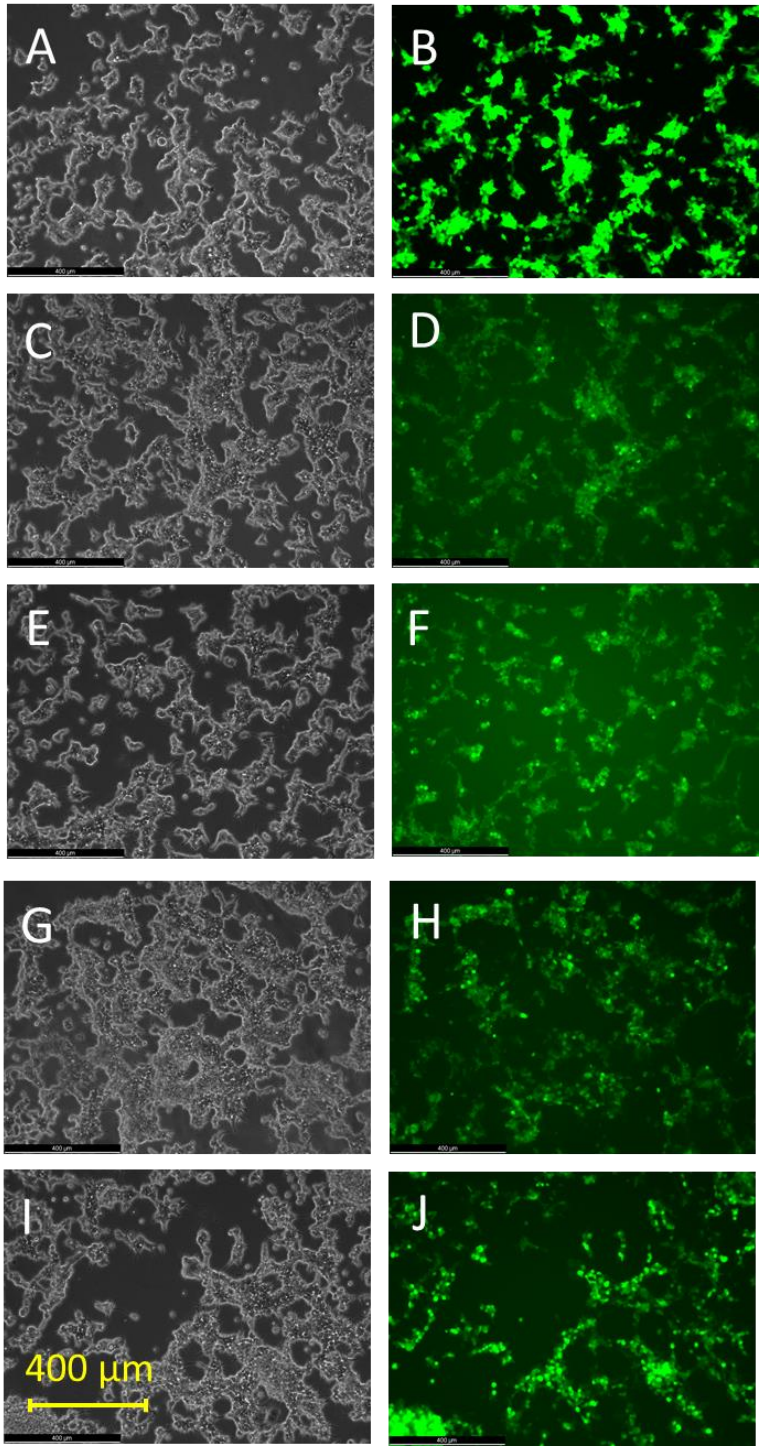


Figure 5.5: Transfection of HEK293T cells with pcDNA3.1(+) WISP1-GFP vectors in 6-well plates. Transfection was done using TransIT-2020 transfection reagent at a 3:1 transfection reagent: DNA ratio. Cells were plated at 500,000 cells/well 24 hours before transfection. Left-hand images are brightfield, right-hand images are GFP fluorescence (509nm) of the same area of cells. Transfected vectors contained: A. and B. GFP. C. and D. Full-length WISP1-GFP. E and F. WISP1 variant 2-GFP. G. and H. WISP1 variant 3-GFP. I. and J. WISP1 variant 4-GFP.

5.6 Western blots of lysates from HEK293T cells expressing WISP-1 splice variants show bands at the correct molecular weight.

Having demonstrated expression using fluorescence microscopy, the next step was to verify that the molecular weight of the proteins being produced were correct. HEK293T cells were chosen, as they can be transfected at very high efficiency and produce large amounts of protein from this transfection, thus maximising the amount of protein of interest produced per cell. HEK293T cell lysates transfected with WISP-1 splice variants were analysed on a western blot and produced band sizes corresponding to the expected sizes of the tagged WISP-1 variants shown in Figure 5.6. Notably, blotting using an anti-WISP-1 primary antibody (Abcam, ab65943) did not detect variant 3 as it was raised against a C-terminal WISP-1 peptide (ESYPDFSEIAN) not present in that variant. The correct sized band (approximately 49kDa) does appear using an anti-GFP western blot. Full-length WISP-1-GFP produces a band at slightly lower than the expected molecular weight of 73kDa.

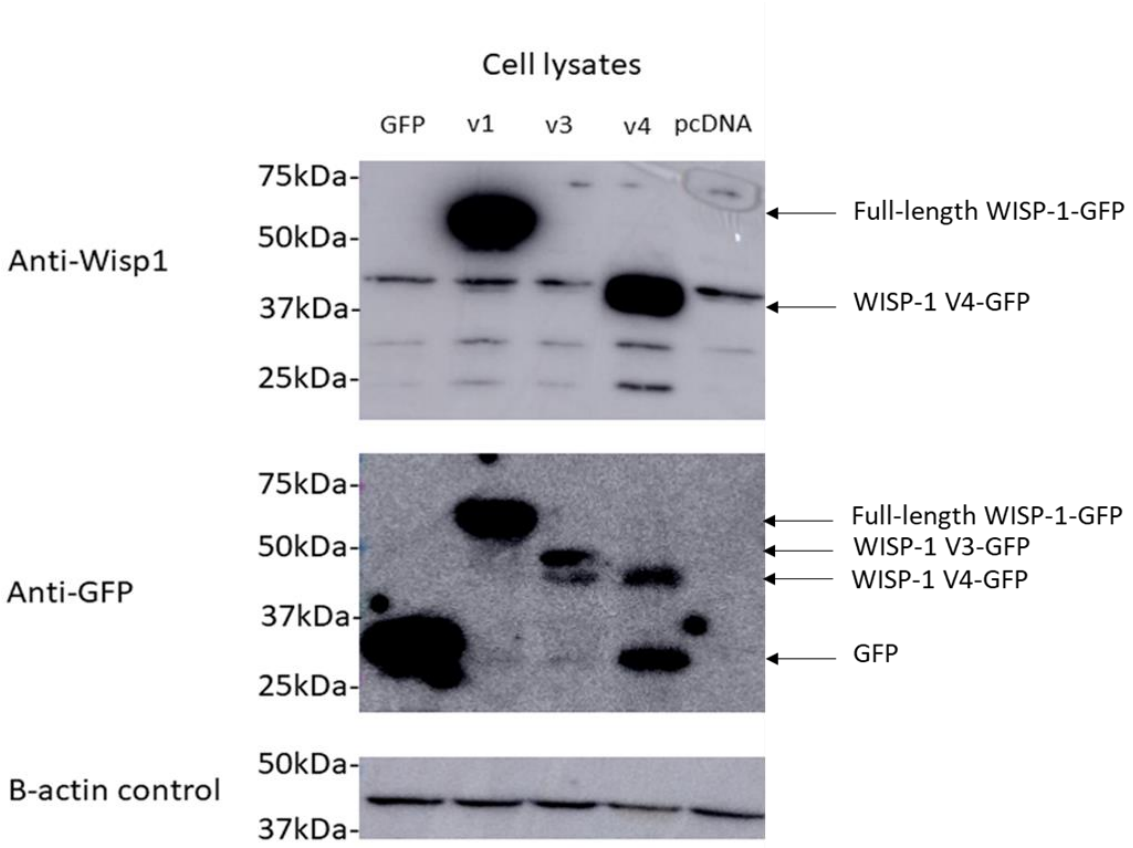


Figure 5.6: Western blot of HEK293T cell lysates prepared 48h after transfection with WISP-1-GFP constructs. 30µg of total protein as measured on a BCA assay was loaded into each lane. V1

is full-length WISP-1, V3 and v4 are variant 3 and 4 respectively. Top: Anti-WISP-1 antibody. Middle: Anti-GFP antibody. Bottom: Anti-beta-actin loading control. Expected molecular weights of tagged splice variants shown are 73kDa, 49kDa, and 46kDa for variants 1, 3 and 4 respectively.

5.7 GFP-trap co-immunoprecipitation can efficiently purify GFP-tagged WISP-1

Having established that GFP-tagged WISP-1 can be transfected into HEK 293T cells, affinity purification was then optimised using this cell line. Immunoprecipitation was done under non-denaturing conditions using Chromotek's GFP-trap immunoprecipitation kit, which uses a high affinity camelid anti-GFP antibody conjugated to agarose beads to allow efficient purification of GFP-tagged proteins and any proteins which interact with them. Full length WISP-1 was chosen to take forward for the initial AP-MS experiments, as it contains all WISP-1 domains, and thus allows for the widest range of possible protein-protein interactions.

Although not hugely relevant as a model for IPF fibroblasts, HEK-293T cells have the great advantage of producing readily detectable amounts of protein. Transfection, cell lysis and co-IP of HEK-293T cells with full-length GFP-tagged WISP-1 showed that the GFP-trap protocol was effective at purifying levels of WISP-1 which are readily detectable by western blot, while efficiently removing other, non-specific proteins from the cell lysates, as demonstrated in Figure 5.7, showing Coomassie brilliant blue protein staining of a western blot gel loaded with the input, flow-though, wash and elution steps of the GFP-trap kit. The flow-through fraction still contained detectable GFP-tagged WISP-1, suggesting that the amount of recombinant protein expressed by HEK293T cells exceeded the capacity of the quantity of GFP-trap recommended for use in the manufacturer's standard protocol. More beads/reaction may have reduced this issue, but the large amount of WISP-1-GFP in the elution fraction was sufficient for subsequent experiments, so this was not attempted.

Figure 5.7: Western blot and Coomassie stain of GFP-trap co-immunoprecipitation of full-length GFP-tagged WISP-1. A. Western blot using anti-GFP antibody of lysates from HEK- 293T cells expressing GFP-tagged full-length WISP-1. Input is cell lysate, flow-through is what remains in the lysate after incubation with the beads, washes 1-3 are taken from washes with bead dilution buffer, and elution is the final elution using the kit elution buffer. Expected molecular weight of WISP-1-GFP is 73kDa. B. Coomassie brilliant blue staining of the gel from A, showing protein staining.

5.8 Mass spectrometry in HEK-293T cells shows WISP-1 interacting with structural cell proteins

Having established that it is possible to affinity purify GFP-tagged full-length WISP-1 from HEK-293T cell lysates, the elution fraction from this experiment was analysed using a mass spectrometer (Orbitrap LC-MS) to identify potential WISP-1 protein-protein interaction partners. For these experiments, lysates prepared from cells expressing GFP alone were used as control. In total, 1369 proteins were positively identified across all samples, including GFP control samples. 660 of these proteins were present in both of the triplicate GFP-tagged WISP-1 samples. Of these, 52 proteins were identified in the GFP alone control samples, and were excluded from downstream analysis as they were deemed to be contaminants. Thus, in total, 608 proteins were identified in this experiment as present in the WISP-1 elution fractions, but not the GFP-only elution fractions. Table 5.1 shows the top 20 proteins identified as interacting with WISP-1 in this study, as ordered by number of peptide spectrum matches. This is the number of peptide spectra (including multiple detections of identical peptides) mapped to each protein and can be used as

an approximation of protein abundance. Notably, multiple tubulin proteins were identified, suggesting a possible interaction of WISP-1 with cell structural proteins. WISP-1 is clearly identified, and is very abundant in these data, showing efficient WISP-1 purification.

Table 5.1: Top 20 proteins identified as interacting with WISP-1 in HEK-293T cells transfected with GFP-tagged full-length WISP-1. Three technical replicates were performed. Proteins shown were identified in all three GFP-tagged WISP-1 elution fractions but were absent in elution fractions taken from HEK-293T cells transfected with a plasmid expressing only GFP, thus eliminating non-specific interactors with the GFP tag. Proteins are ordered by number of peptide spectrum matches.

Description	Gene Name	Number of peptide spectrum matches
DNA-dependent protein kinase catalytic subunit	PRKDC	107
Tubulin beta chain	TUBB	106
Tubulin beta-4B chain	TUBB4B	100
Tubulin beta-2A chain	TUBB2A	83
Tubulin beta-2B chain	TUBB2B	82
Tubulin beta-4A chain	TUBB4A	80
WNT1-inducible-signaling pathway protein 1	WISP1	72
Cytoplasmic dynein 1 heavy chain 1	DYNC1H1	69
Tubulin alpha-1B chain	TUBA1B	59
Tubulin alpha chain	TUBA1C	53
CAD protein	CAD	45
Heat shock protein HSP 90-beta	HSP90AB1	30
Tubulin beta-6 chain	TUBB6	27
Coatomer subunit alpha	COPA	25
Fatty acid synthase	FASN	22
eIF-2-alpha kinase activator GCN1	GCN1	22
Sodium/potassium-transporting ATPase subunit alpha-1	ATP1A1	20
Heat shock protein HSP 90-alpha	HSP90AA1	20
D-3-phphoglycerate dehydrogenase	PHGDH	19
E3 ubiquitin-protein ligase UBR4	UBR4	19

5.9 WISP-1 may be involved in telomerase regulation

Given the large number of potential interaction partners identified by this initial study, gene ontology analysis was used to identify cellular processes which these interaction partners may have in common. All the identified protein interaction partners, as well as WISP-1 itself, were input into the GOrilla gene ontology database as a ranked list of genes.

Figure 5.8 shows the cellular localisation and cellular process GO terms associated with these WISP-1 interaction partners. The treemap on the right includes a large number of gene ontology terms related to the positive regulation of protein localisation to telomeres. This is largely due to protein interactions identified between WISP-1 and the chaperonin containing tcp1 (CCT) complex, a protein complex of molecular chaperones. CCT is upregulated in many cancers and is responsible for assisting the folding of any cancer-associated proteins. CCT is a protein complex consisting of two stacked rings of 8 similar protein subunits. It has been implicated in catalysing the folding of many key signalling proteins, including KRAS(236). It is also an important regulator of protein trafficking to telomerase.

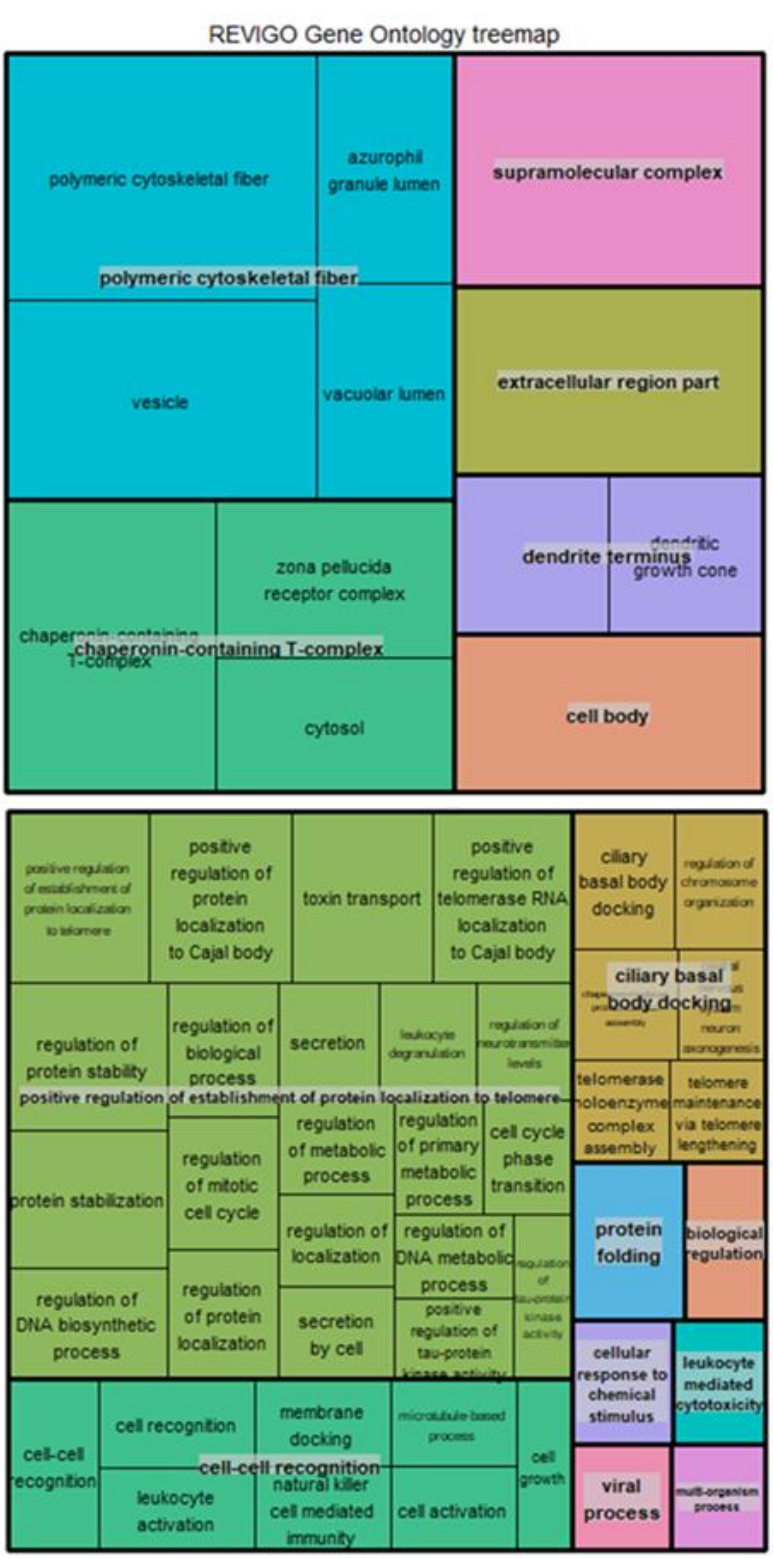


Figure 5.8: Treemaps showing gene ontology terms identified as relevant to WISP-1 protein interaction partners identified in HEK-293T cells. Top – Cellular compartment GO terms. Bottom – Cellular process related GO terms. Box size is proportional to the p-values associated with the GO terms, with a lower p-value giving a larger box.

5.10 Transfection of MRC-5 cells with a GFP containing expression constructs shows strong expression levels.

Having established that WISP-1 protein interaction partners could be identified in HEK-293T cells, a more appropriate model for IPF was sought. As WISP-1 is largely expressed in lung fibroblasts in IPF patients, a lung fibroblast cell line, MRC-5, was selected. These cells are fast growing, while also being a more appropriate model system for looking at the function role of WISP-1 in IPF than kidney cells.

Initially transfection was performed using a pcDNA3.1(+) plasmid only containing the GFP gene to optimise the transfection conditions. Figure 5.9 shows GFP fluorescence microscopy of MRC-5 cells 24 hours post-transfection with three different transfection reagents: lipofectamine 2000, transIT 2020 and transIT X2. Of these, transIT X2 at a 3:1 transfection reagent:DNA volume:mass ratio gave the best results, with low cellular toxicity and high transfection efficiency . Transfection efficiency was approximately 70%, although there was a large variation in the intensity of fluorescence exhibited by different cells (n.b. Figure 5.9F does not show fluorescence from all transfected cells as higher exposure would overexpose GFP-high cells).

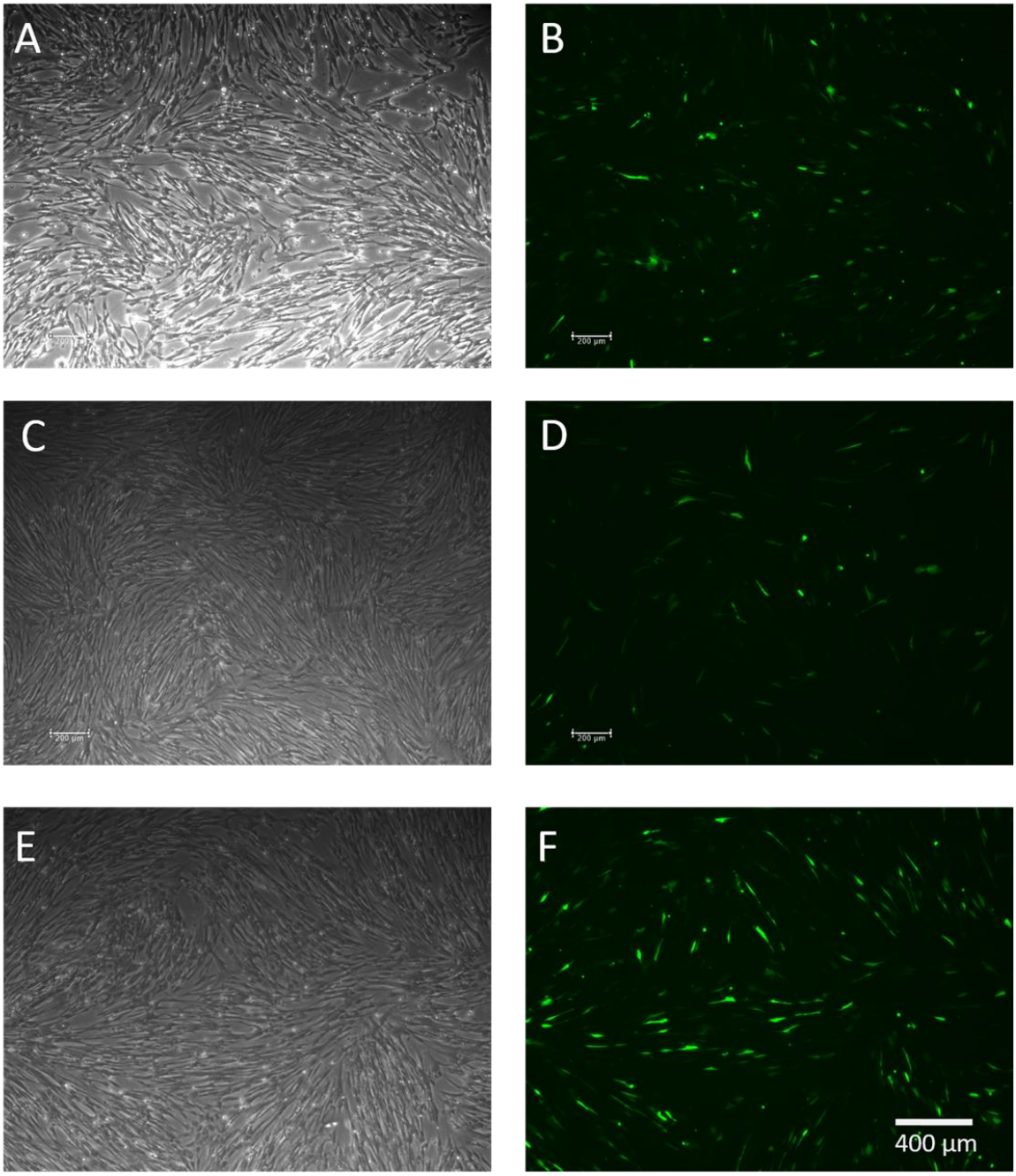


Figure 5.9: Optimisation of MRC-5 cell transfection with the pcDNA3.1(+)GFP construct. A and B - Phase-contrast and 509nm fluorescence images of MRC-5 cells 24 hours post-transfection with lipofectamine 2000 at a 2:1 transfection reagent: DNA ratio (10µl lipofectamine, 5µg DNA in one well of a 6 well plate). B and C - Phase contrast and fluorescence images of MRC-5 cells transfected using TransIT 2020 at a 3:1 Transfection reagent: DNA ratio. D and E - Phase contrast and fluorescence images of MRC-5 cells transfected using TransIT X2 at a 3:1 Transfection reagent:DNA ratio (7.5µl transfection reagent, 2.5µg DNA per well in a 6-well plate).

5.11 GFP-tagged WISP-1 expression constructs show reduced, but detectable expression levels in MRC-5 fibroblasts.

Having optimised transfection conditions in MRC-5 cells and achieved a high efficiency of transfection with GFP-only containing plasmid, MRC-5 cells were then transfected with a plasmid containing WISP1-GFP. Figure 5.10 shows GFP fluorescence of MRC-5 cells transfected with full-length WISP1-GFP expression construct, at a 3:1 Transit-X2: DNA ratio. Transfection efficiency in terms of number of cells showing some fluorescence was comparable to the GFP transfected MRC-5 cells, however, similar to the HEK293T cells, there was a substantial decrease in fluorescence intensity for WISP-1-GFP compared to GFP alone. This could be explained by WISP-1-GFP having a higher degradation rate than GFP, or by it being exported the cells, or WISP-1-GFP mRNA being degraded at a higher rate. Although lower, this amount of WISP-1 expression proved sufficient to reliably immunoprecipitate WISP-1-GFP from MRC-5 lysates. Cells were harvested at 48 hours post-transfection. Although there was a small decrease in cell numbers by this timepoint compared to at 24 hours, this was more than compensated for by higher WISP-1-GFP expression.

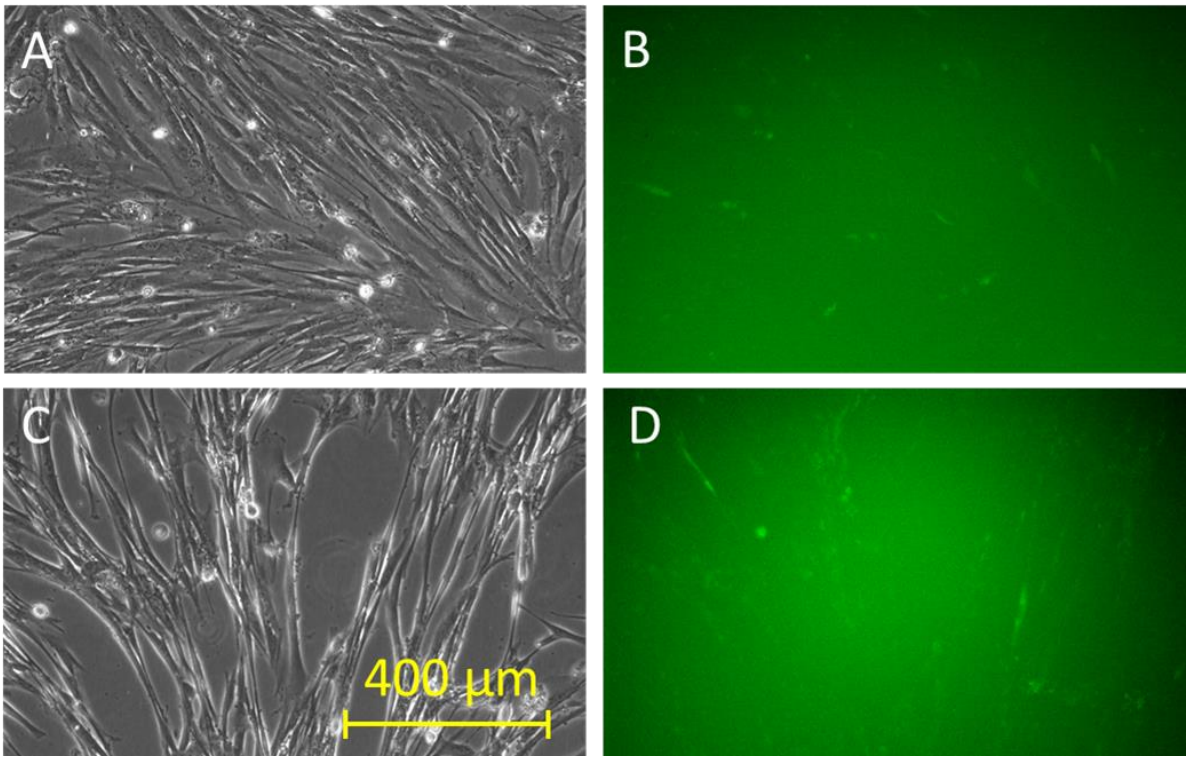


Figure 5.10: Fluorescent microscopy of MRC-5 cells transfected with GFP-tagged full-length WISP1 pcDNA3.1(+) expression construct in 6-well plates using Transit-X2 at a 3:1 reagent: DNA ratio. Left-hand images are phase-contrast, right-hand images are A. and B. Images taken 24 hours post-transfection. C. and D. Images taken at 48 hours post-transfection, prior to harvesting.

5.12 Immunoprecipitation of GFP-tagged WISP-1 from MRC-5 lung fibroblasts shows efficient purification.

Once it was clear that GFP-tagged WISP-1 could be efficiently transfected into MRC-5 lung fibroblasts, GFP-trap affinity purification was used to purify GFP-tagged WISP-1 from transfected MRC-5 cell lysates. Figure 5.11 shows a western blot using an anti-WISP-1 antibody of input, flow-through and elution fractions from this immunoprecipitation. There is a clear, substantial depletion of GFP-tagged WISP-1 in the flow-through compared to the WISP-1 input, and GFP-tagged WISP-1 is readily detectable in the elution fraction. The blot shown in Figure 5.11 shows elution from the GFP-trap beads using an acidic glycine elution buffer. This was required for preparation of the elution fractions for mass spectrometry. However, it is clear from the density of the bands that the amount of WISP-1 present in the elution fraction is less than the amount removed from the input fraction, suggesting that this elution method is insufficient to completely

remove all the bound GFP-tagged WISP1 from the beads. For this blot, input and flow-through fractions were identical except for protein bound to beads, so are directly comparable. Elution fractions should result in a higher concentration (approximately 10-fold) of the bound protein. The observed reduction in WISP-1-GFP concentration between the input and elution fractions, even though proteins were eluted in a smaller volume than the input, is evidence that a large amount of WISP-1-GFP is retained on the beads using this elution protocol. Although this protocol was required for mass spectrometry, subsequent Western blots to validate interaction partners (Figure 5.13) were performed on elution fractions from beads heated to 95°C in reducing SDS sample buffer, which yielded substantially greater amounts of protein.

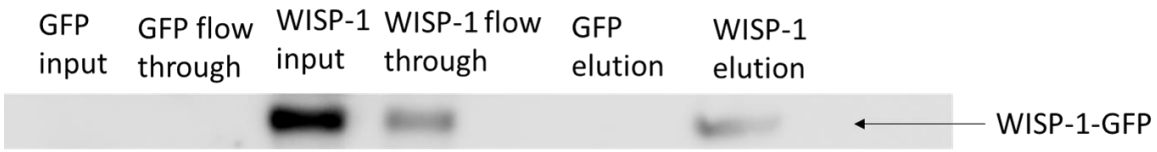


Figure 5.11: Western blot of GFP-trap input, flow-through and elution fractions for MRC-5 cells transfected with constructs expressing GFP, and GFP-tagged WISP-1. Input is diluted lysate before incubation with anti-GFP coated beads, flow-through is diluted lysate after bead incubation, and elution is the elution fraction after treating the beads with 50µl of an acidic glycine solution. Equal volumes of each fraction (10µl) were loaded in each well. Input and flow-through are thus directly comparable, while elution fractions should result in a higher concentration

5.13 Mass spectrometry shows WISP-1 interacts with basic fibroblast growth factor (FGF2) and mitochondrial cell survival proteins

Having generated GFP-tagged WISP-1 expression constructs, in order to investigate the functional importance of WISP-1, mass spectrometry was done on elution fractions from GFP-trap beads prepared using lysates from MRC-5 cells transfected with WISP1-GFP, using lysates from the cells transfected with a GFP only as a control for non-specific binding. Expression of both proteins was verified by fluorescence microscopy before cells were lysed. Lysates were subjected to GFP-trap immunoprecipitation before proteins were eluted in acidic (pH 3) glycine buffer. Equal quantities of protein from these elution fractions were then analysed by mass spectrometry, using the

Orbitrap mass spectrometer to identify WISP1 protein interaction partners. In total, 313 proteins were identified in at least one sample from the WISP-1-GFP elution fractions. Of these proteins, 236 were present in at least one GFP sample. These were removed from the data as contaminants, leaving only 77 proteins identified in WISP-1-GFP elution fractions. Of these, 12 proteins were identified in both WISP-1-GFP containing samples. The 12 Interaction partners present in both GFP-tagged WISP1 samples, but not present in the GFP samples are listed in Table 5.2. As with the HEK293T elution fractions, gene ontology analysis was attempted, but too few proteins were identified to discern any significant GO terms.

Of the proteins identified, it is likely that trypsin is a contaminant from the mass spectrometry processing, as trypsin was used to lyse proteins into tryptic peptides during the mass spectrometry sample processing. It is unclear why trypsin was not identified as a contaminant in the GFP-only elution fractions, but it may be due to an interaction between WISP-1 and trypsin which does not occur with GFP. Of the other proteins identified, several are of considerable interest in the context of IPF. Basic fibroblast growth factor (FGF2) is a protein associated with fibroblast proliferation(237). It is mitogenic, and associated with cellular survival(238). Prohibitin (PHB) and voltage dependent anion-selective channel protein 1 (VDAC1) are both mitochondrial cell survival proteins which have been associated with increased resistance to apoptosis. Prohibitin is an inner mitochondrial membrane-bound protein which is highly conserved in all eukaryotes(239). It likely functions to stabilise mitochondrial structure during formation, but has also been identified as a regulator of cellular proliferation. VDAC1, likewise, is a mitochondrial ion channel protein. It transports calcium ions across the mitochondrial membrane. It has been shown to interact with the Bcl family of pro-apoptotic proteins, thus suggesting it plays a role in regulating cellular survival and proliferation(240,241).

In summary, in MRC-5 lung fibroblasts, WISP-1 interacts with a growth factor known to promote fibroblast proliferation, as well as two distinct mitochondrial proteins known to promote cell survival proteins. This suggests WISP-1 may play a role in promoting the fibroblast proliferation and survival characteristics of fibroblastic foci in IPF.

Table 5.2: Protein interaction partners of full-length WISP-1 in MRC-5 lung fibroblasts. Data shown is from two technical replicates, showing proteins present in both samples and not in elution fractions from GFP-transfected MRC-5 cells. Proteins are listed in order of their number of peptide spectrum matches. Proteins of specific interest are highlighted in yellow.

Name	Gene	Number of peptide spectrum matches
WNT1-inducible-signaling pathway protein 1	WISP1	18
Trypsin-1	PRSS1	9
HLA class I histocompatibility antigen, Cw-6 alpha chain	HLA-C	4
Dermcidin	DCD	2
Fibroblast growth factor	FGF2	2
Calreticulin	CALR	3
Voltage-dependent anion-selective channel protein 1	VDAC1	3
Rho GDP-dissociation inhibitor 1	ARHGDIA	2
Protein POF1B	POF1B	2
Proteasome subunit alpha type-1	PSMA1	2
40S ribosomal protein S27	RPS27L	2
Prohibitin	PHB	2

5.14 Mass spectrometry of WISP-1 elution fractions detected additional proteins in individual samples.

While the proteins described in Table 5.2 are the most robustly identified, as they were detected in both GFP-tagged WISP-1 samples, there were several proteins only identified in a single sample which may be relevant to the role of WISP-1 in IPF. Although their identity as genuine protein interaction partners is less robust, there are several proteins present which, if they are actual interaction partners, could have important implications for WISP-1's role in IPF pathogenesis. Table 5.3 shows a curated subset of these interactors.

Most notably, the collagen crosslinking enzyme PLOD3 is highly upregulated in this dataset. PLOD3 has been linked with fibrosis. Knockdown of PLOD3 in the A549 alveolar epithelial cell line resulted in lower collagen crosslinking activity, as well as a decrease in TGF-β and Wnt signalling. Overexpression of PLOD3 led to an increase in collagen deposition and increased activity of both

profibrotic signalling pathways(242). The related crosslinking enzymes, PLOD2 or LOXL2 have also been linked to fibrosis through increasing bone-type collagen crosslinking in IPF tissue, leading to highly stiffened, collagen fibrils(65).

Other proteins identified include Metallothionein-1G. This is a protein associated with metal ion sensing but is also important in sensing of reactive oxygen species. The metallothionein proteins are also strongly upregulated in the PLIN2+ fibroblasts identified in the single-cell RNA-seq dataset presented in chapter 3(58) (Table 3.12).Other interaction partners include zinc finger protein 410, which is a transcription factor which has been associated with transcriptional activation of extracellular matrix genes (including MMP1) in fibroblast senescence(243).

There are also several proteins associated with intracellular signalling, including calmodulin-like protein 3, E3 ubiquitin-protein ligase (SIAH1), and two Ras-related proteins, all suggesting a role for WISP-1 in altering cell signal transduction. 14-3-3 protein zeta/delta (YWHAZ) is a pro-proliferative protein that is important for signal transduction and regulation of apoptosis as well.

Apolipoprotein L2 is a lipid-binding protein, whose interaction with WISP-1 may implicate WISP-1 in playing a role as an adipokine, a role which it has previously been associated with in other diseases.

Finally, heat shock 70kDa protein 6 (HSPA6) is a heat shock protein associated with the cellular stress response, and hypoxia up-regulated protein 1 (HYOU1), which is involved in hypoxia signalling, show WISP-1 interacting with proteins induced in response to cellular stresses. In particular, the identification of HYOU1, a protein which provides a protective effect to cells in hypoxic condition as interacting with WISP-1, suggests that the induction of WISP-1 by hypoxia, demonstrated in the previous results chapter of this thesis, may implicate WISP-1 as being important for the hypoxia response of lung fibroblasts in IPF.

Although these interaction partners are necessarily less certain than the ones identified in both samples above, they identify diverse roles for WISP-1 in cellular signalling, stress and hypoxia response, regulation of lipid binding, oxidative stress sensing, cellular survival, and extracellular matrix production and collagen crosslinking.

Table 5.3: A subset of WISP-1 interactors identified in only one (of two) elution fraction from MRC-5 cells. Ordered alphabetically.

Protein Name	Gene
14-3-3 protein zeta/delta	YWHAZ
Apolipoprotein L2	APOL2
Calmodulin-like protein 3	CALML3
E3 ubiquitin-protein ligase SIAH1	SIAH1
Heat shock 70 kDa protein 6	HSPA6
Hypoxia up-regulated protein 1	HYOU1
Metallothionein-1G	MT1G
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	PLOD3
Ras-related protein R-Ras2	RRAS2
Ras-related protein Rab-10	RAB10
Zinc finger protein 410	ZNF410

5.15 Three WISP-1 protein interaction partners identified are common to both MRC-5 and HEK-293T lists

Although there is a substantial difference in the list of proteins identified between both cell lines, there were three proteins identified as being common interaction partners between both cell lines. These were VDAC1, prohibitin, and dermicidin (DCD). Both VDAC1 and prohibitin have previously been discussed in the context of mitochondrial cell survival proteins which may be important for regulating apoptosis and cell proliferation, while dermicidin is an antimicrobial peptide important for innate immunity.

5.16 VDAC1, PHB and FGF2 are all detectable in MRC-5 cell lysates and flow-through fractions

Havin identified WISP-1 interaction partners of interest by mass spectrometry, western blot analysis was performed on the input and flow-through fractions of this Co-IP experiment to confirm that these proteins were expressed in IPF cells. The proteins PHB, VDAC1 and FGF2 were selected for further analysis – the former two because they were also found to interact with WISP-1 in HEK293T cells, as well as their potential role in cell proliferation, survival and apoptosis,

the latter because it is a potent mitogen associated with cellular proliferation, including proliferation in fibroblasts and osteoblasts.(237,244)

Figure 5.12 shows western blots of these proteins, in the input and flow-through fractions from the GFP-trap Co-IP. All three proteins were detected. Interestingly, although the amount of VDAC1 is unchanged, the amounts of both PHB and FGF2 are increased in the WISP-1-GFP transfected cells, suggesting that WISP-1 overexpression may drive the production of these proteins. Three separate bands were observed for FGF2, likely corresponding to low and high molecular weight FGF2. These isoforms have been identified as cytoplasmic and nuclear, respectively, and may play different roles in mediating FGF2 signalling(245).

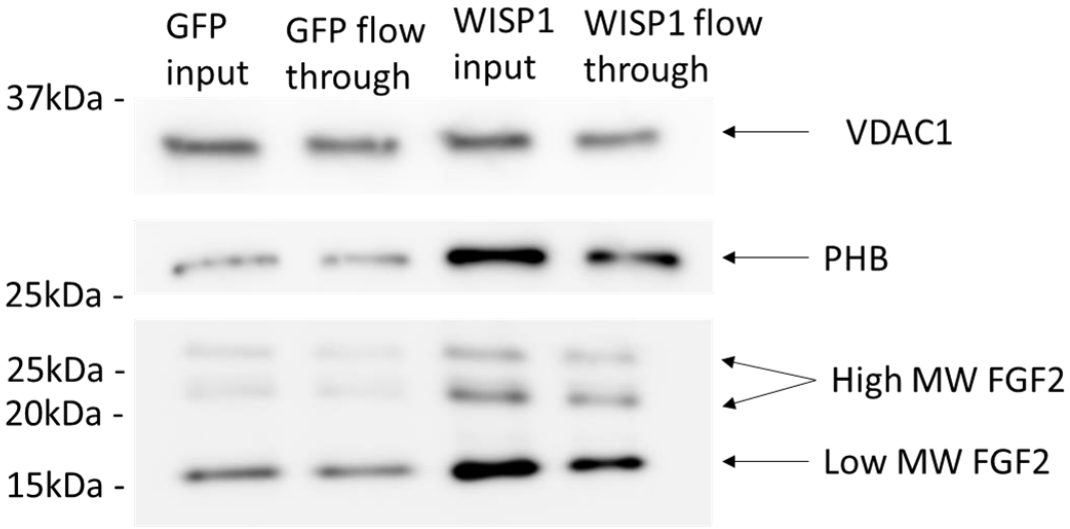


Figure 5.12: Western blots of WISP-1 interaction partners in MRC-5 lysates transfected with GFP alone or GFP-tagged WISP-1. Input is diluted cell lysate, flow through is cell lysate after incubation with GFP-trap beads. Loading amounts between input fractions were standardised using a BCA assay. The same volume of input and flow-through fractions were loaded for each transfection to identify any drop in protein level due to WISP-1 binding.

5.17 Western blot analysis confirms the interaction of WISP-1 with FGF2 and VDAC1

In order to confirm interactions of WISP-1 with FGF2, VDAC1 and prohibitin, western blot analysis was performed on the elution fractions from GFP-Trap beads after extraction of binding proteins from lysates of GFP-tagged WISP-1 expressing MRC5 fibroblasts. Elution fractions from cells were standardised to enable equivalent loading on the gels.

Figure 5.13 shows western blot analysis of the elution fractions. Blotting with anti-GFP confirmed similar loading of samples containing WISP-1 tagged GFP and GFP alone; while both FGF2 and VDAC1 were readily detectable in the WISP-1 GFP elution fraction, they were absent in the GFP-only fraction. A western blot using an anti-prohibitin antibody was also performed, but no prohibitin was detectable in the elution fractions using this antibody.

FGF2 has several different isoforms, including high and low molecular weight versions. This analysis detected bands at approximately 18kDa and 24kDa, most likely corresponding to cytoplasmic and nuclear FGF2 respectively(245). There is a fainter band above the 24kDa band highlighted – likely corresponding to the highest molecular weight band identified in Figure 5.12. It is clear that there is enrichment for high-MW FGF2 compared to low-MW in this elution fraction compared to their relative band densities in the input and flow-through fractions (Figure 5.12). This suggests that WISP-1 may preferentially interact with high-MW FGF2 in these cells. As high-MW FGF2 has been shown to localise to the nucleus(245), it raises the possibility that WISP-1 may play a role in nuclear signalling. This may explain the partial nuclear localisation of WISP-1-GFP identified in Figure 5.4.

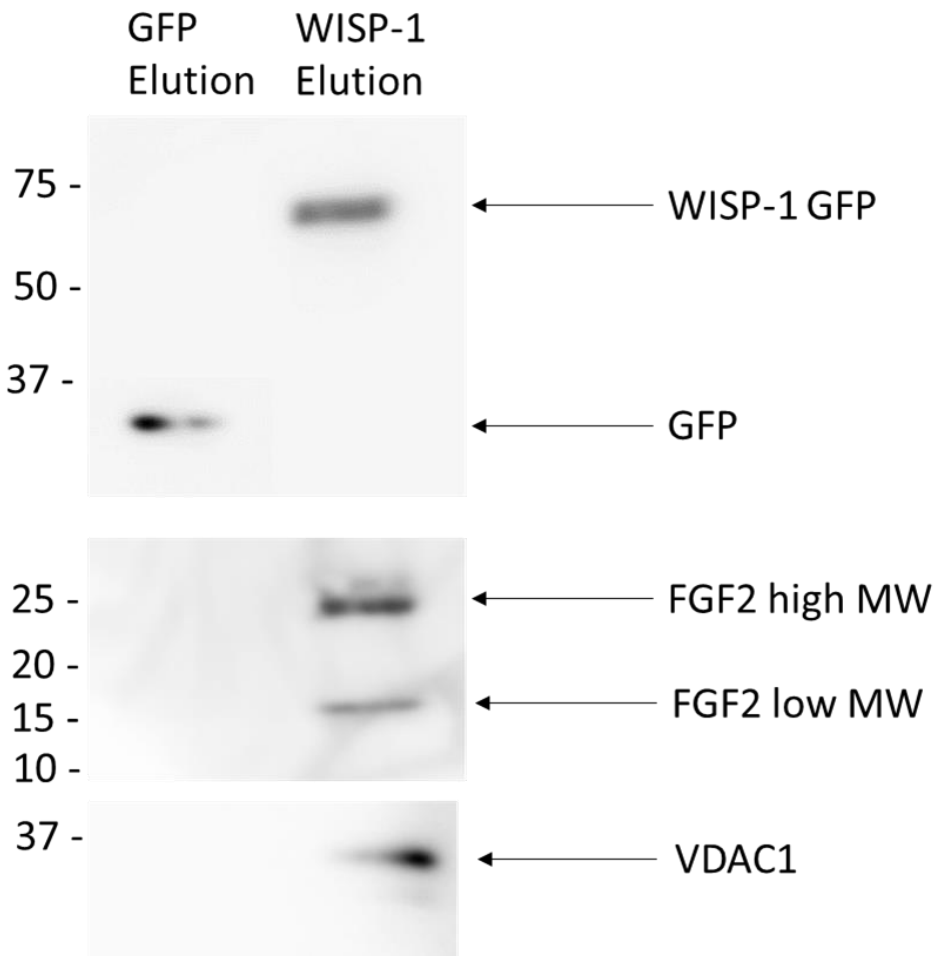


Figure 5.13: Western blot analysis of GFP-trap elution fractions from MRC-5 cells transfected with GFP alone or GFP-tagged WISP-1. Top – western blot using an anti-GFP antibody, showing bands corresponding to GFP-alone (MW…) and GFP-tagged WISP-1 (MW….). Middle – western blot of the same samples using an anti-FGF2 antibody, with bands corresponding to high and low molecular weight FGF2 only in the GFP-tagged WISP-1 sample. Bottom – western blot of the same samples using an anti-VDAC1 antibody.

5.18 A protein BLAST search reveals that WISP1 has significant sequence homology to CTGF.

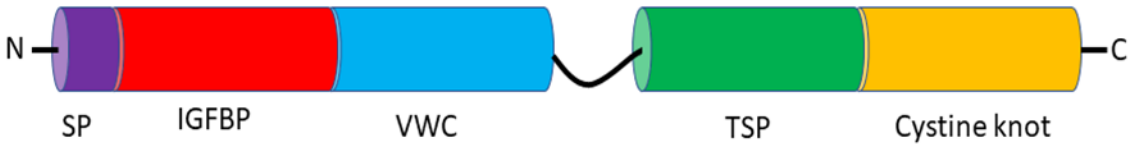
Notably, FGF2 has been shown to interact with another member of the CCN protein family, connective tissue growth factor, or CTGF, also known as CCN2(183). This is a protein which is well established as being important for cellular proliferation and is a key downstream effector of TGFB. CTGF also has an identical domain structure to WISP-1 (Figure 5.14B), and significant protein homology . The study which established this interaction identified the CTGF-FGF2

interaction as contributing to survival, proliferation and activation of chondrocytes, a cartilage-producing subtype of fibroblast. This study localised the CTGF-FGF2 interaction to the thromopospondin response element domain of CTGF(183).

The results of a protein BLAST alignment are shown in Figure 5.14A. The TSP1 domain of WISP-1 has a strong homology to the corresponding domain in CTGF/CCN2, with 52% of amino acids being identical to the ones in the same position in both proteins, and 74% being positive. Positives are amino acids which have similar chemical properties in both domains, even though they are not identical, such as isoleucine (I) and Leucine (L)), and so the overall function or structure of that part of the protein is conserved.

The TSP1 domains of WISP-1 and CTGF are both very similar, increasing the likelihood of them having similar protein binding properties.

A



B

Score	Expect	Method	Identities	Positives	Gaps
67.4 bits(163)	9e-23	Compositional matrix adjust.	28/54(52%)	40/54(74%)	0/54(0%)

```
Query  1   NCIAYTSPWSPCSTSCGLGVSTRISNVNAQCWPEQESRLCNLRPCDVDIHTLIK  54
           NC+  T+ WS CS +CG+G+STR++N NA C  E++SRLC +RPC+ D+   IK
Sbjct  1   NCLVQTTEWSACSKTCGMGISTRVTNDNASCRLEKQSRLCMVRPCEADLEENIK  54
```

Figure 5.14: A. Domain structure of CCN proteins, including WISP-1 and CTGF. B. Results of protein basic local alignment search tool (BLAST) search using the amino acid sequence of the thrombospondin response element (TSP) domain of WISP-1 as the query, and the sequence of the TSP domain of CTGF/CCN2 as the subject. The middle line shows homologous amino acids between the two domains, with "+" signs denoting positive amino acid substitutions likely to conserve function.

5.19 An ELISA-based protein binding assay probing the WISP-1 – FGF2 interaction *in vitro* shows WISP-1 binding to FGF2.

Having established that WISP-1 interacts with FGF2 in the tagged overexpression model used to identify WISP-1 protein interaction partners in cultured lung fibroblasts, and having confirmed that interaction using western blot analysis using an anti-FGF2 antibody, an enzyme-linked immuno-sorbent assay (ELISA) was used to quantify the concentration-dependence of this interaction. In order to test this, a modified binding assay based on the method utilised for analysis of the interaction between CTGF and FGF2 was used, except using recombinant WISP-1 rather than CTGF(183). This assay involves examination of the binding of a fixed concentration of WISP-1 (25nM) to increasing concentrations of FGF2. As recombinant high-MW FGF2 was not readily available, low-MW FGF2 was used Figure 5.15 shows the results of a preliminary binding assay. Binding of WISP-1 to FGF2 was readily detectable and showed a clear dose-response, with maximal binding occurring at concentrations of FGF2 in excess of 100nM.

Another key question based on the identification of CTGF as an FGF2 interaction partner is whether CTGF and WISP-1 compete for binding at the same location on the FGF2 molecule. To test this, a preliminary experiment was performed examining the effect of CTGF on WISP-1 binding to FGF2. Two limiting concentrations (22nM and 44nM) of FGF2 were selected and binding of WISP-1 (25nM) was assessed in the absence or presence of CTGF (25nM). At these FGF2 concentrations, CTGF caused a small reduction in WISP-1 binding (Figure 5.15B), suggesting that WISP-1 and CTGF may compete for the same or overlapping sites on the FGF2 protein. However, further binding assays will be required to validate these findings.

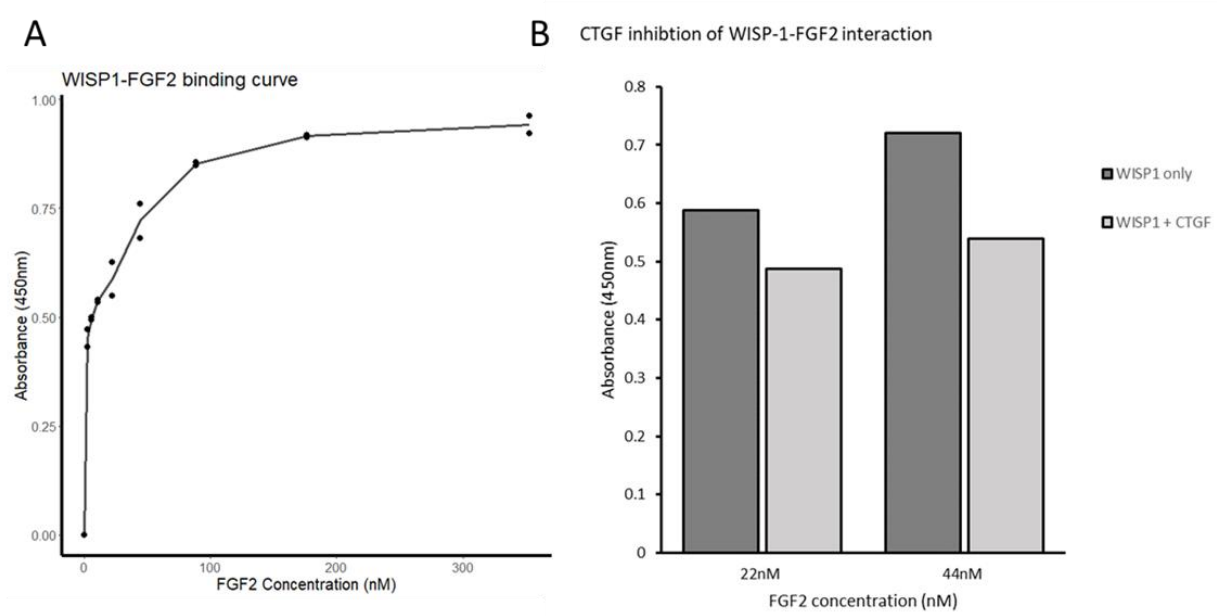


Figure 5.15: The interaction between FGF2 and WISP-1. Recombinant WISP-1 at a concentration of 25nM was added to wells of a 96-well plate coated with serial dilutions of recombinant FGF2. WISP-1 binding was measured using a biotinylated anti WISP-1 antibody with streptavidin-HRP enabling detection using H2O2 with TMB as chromogen. Absorbance values were corrected for any nonspecific binding of WISP-1 in the absence of FGF2. Measurements were performed in duplicate and the mean of each duplicate is presented (N=1 experiment). B. Bar chart showing inhibition of the WISP-1-FGF2 interaction by CTGF. Equimolar concentrations of WISP-1 and CTGF at 25nM were added to wells coated with FGF2 at 22nM and 44nM concentrations. Absorbance values are compared to values from wells with only WISP-1 added. Bars show the mean of duplicate wells (N=1 experiment)

5.20 The expression pattern of interaction partners suggests WISP-1 undergoes cell type-specific protein interactions

Having identified several WISP-1 protein interaction partners in MRC-5 lung fibroblasts, their gene expression in the mesenchymal cells from the Banovich/Kropski single cell RNAseq dataset was examined. Figure 5.16 shows the expression pattern of VDAC1, PHB1, FGF2 and CTGF compared to the pattern of WISP1 expression in this dataset. Notably, VDAC1 and WISP1 show similar expression patterns, with both enriched in profibrotic myofibroblasts and HAS1-high fibroblasts. CTGF also shows this pattern, but is more generally enriched in myofibroblasts. However, FGF2 (and to a lesser extent PHB1) expression is localised largely to the Has1-high fibroblasts and PLIN2+ fibroblasts. HAS1-high cells are therefore likely to be the site of interaction for WISP-1 and

FGF2, which may have important implications for WISP-1 functionality (see Final Discussion). In summary, mesenchymal cells isolated from IPF lung tissue express both WISP-1 and its interaction partners.

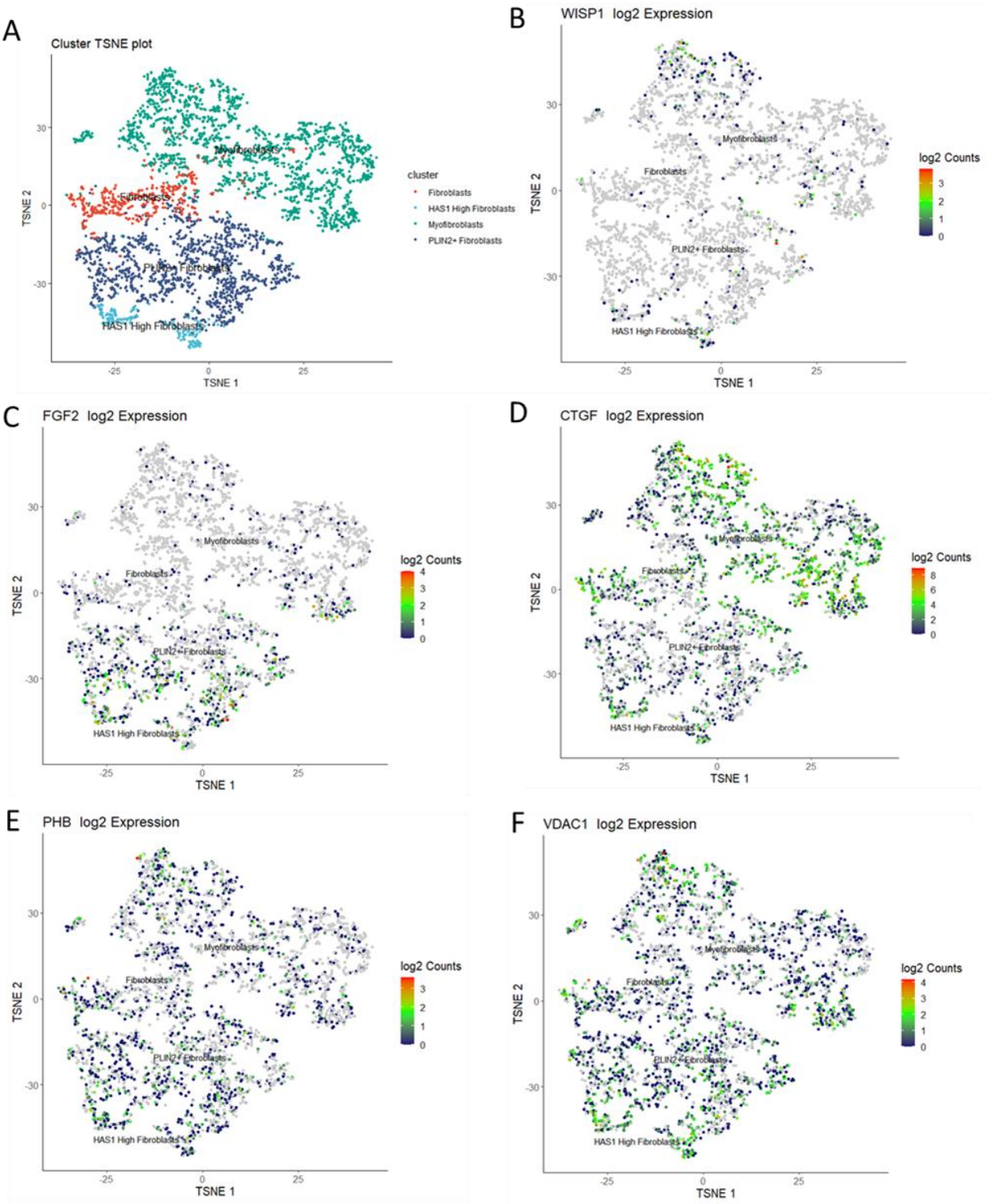


Figure 5.16: t-SNE plots showing the expression pattern of genes coding for WISP-1 and its interaction partners, as well as CTGF, in mesenchymal cells reported in the Banovich/Kropski

single cell RNA-seq dataset. Colour represents log_2(Counts). A. Cluster plot showing different cell types. B. WISP1 expression. C. FGF2 expression. D. CTGF expression. E. PHB expression. F. VDAC1 expression.

5.21 Conclusion

The research presented in this chapter shows that it is possible to overexpress WISP1 variants in various different cell types, and that WISP-1 can be expressed with a GFP tag. This GFP-tagged WISP-1 can be purified using a GFP-trap antibody-based immunoprecipitation protocol, and interaction partners of WISP-1 can be identified in at least two different cell types using mass spectrometry and some of these interactions were confirmed in other independent assays.

This research has identified WISP-1 as having very different protein interaction partners in the two different cell types assayed here. WISP-1 transfection into HEK-293T cells showed a range of protein interaction partners, including proteins associated with telomerase activity and cellular survival, as well as stress response proteins, pointing to a possible role of WISP-1 as a survival factor by modifying activity of telomerase enzymes. However, the protein-protein interactions identified in MRC-5 lung fibroblasts were substantially different to those in HEK-293T cells, although there were some common interaction partners, notably VDAC1 and prohibitin. As MRC-5 cells are cultured lung fibroblasts, they are likely more physiologically relevant to IPF than HEK-293T cells. From an IPF perspective, the interaction with FGF2 is likely to be highly relevant, as it is both a protein that is important for fibroblast proliferation and extracellular matrix production, and its receptors, FGFR-1-4, are targeted by nintedanib, one of two licensed therapies for IPF. Furthermore, FGF2 is known to interact with CTGF/CCN2, a member of the CCN protein family which has an identical domain structure to WISP-1. This research has confirmed this interaction, both by western blot analysis, and by using and ELISA based protein binding assay. The research presented here also identified two mitochondrial proteins previously associated with cell survival as WISP-1 interaction partners, suggesting that WISP-1 may play a role in the increased resistance to apoptosis and cellular proliferation seen in IPF.

5.22 Chapter Discussion

This chapter has focussed firstly on developing a robust methodology to perform co-immunoprecipitation studies on WISP-1, and secondly on identifying WISP-1 interaction partners as a way of generating information on how WISP-1 functions. Members of the CCN protein family are excellent targets for this kind of analysis, as they mediate most of their function through their

interactions with other proteins(121,246). This is in contrast to most other protein types associated with the control of cellular signalling, which may either add post-translational modifications to proteins, or directly bind nucleic acids(98).

The choice to use a tagged overexpression system to perform this analysis, as opposed to attempt to use antibodies against WISP-1 to purify the natively expressed protein, was made for several reasons. Firstly, there is substantial loss of the target protein in any immunoprecipitation assay, and this yield loss is compounded when studying a low-expressing protein such as WISP-1. Also, while there are many commercially available antibodies which bind WISP-1, their binding affinity and non-specific, off target activity (for example, binding other proteins containing similar binding epitopes) are likely to be inferior to highly specific and high-affinity anti-GFP antibodies, such as those in the GFP-trap kit used for the immunoprecipitation here. The GFP-trap kit also has pre-developed, well-optimised protocols for GFP protein purification, reducing the requirement for optimisation of an in-house immunoprecipitation protocol. While overexpression of a target protein is inherently non-physiological, it has the advantage of increasing the sensitivity of a purification assay to less high affinity binding partners which may be undetectable with native protein, but still physiologically relevant. As stated above, due to the low physiological expression levels of WISP-1, an overexpression approach was chosen.

Fluorescence intensities of different splice variants and GFP alone, shown in Figure 5.4, Figure 5.5 and Figure 5.10, are not directly comparable, as different exposures were required to visualise the differing fluorescence intensities of the different splice variants. Fluorescence intensities of GFP-tagged WISP-1 variant transfected cells were broadly similar. However, the fluorescence intensity of GFP alone was considerably higher for all cell lines, with exposures adjusted accordingly. The latter two figures can be used to compare transfection efficiency between different vectors, however, in terms of the number of cells which showed some fluorescence. In this capacity, the different constructs were similar in terms of their overall transfection efficiency.

There are many potential protein tags suitable for a variety of purification techniques, such as nickel column purification for hexahistidine tags, as well as a variety of antibody based peptide affinity methods, of which the FLAG tag is the most well-established(247). The choice of GFP as a tag was due to the availability of high-affinity kits such as GFP-trap which are capable of efficiently purifying GFP and GFP-tagged proteins, as well the ability of GFP to fluoresce, allowing a straightforward assessment of transfection efficiency by observing GFP fluorescence under a microscope. GFP fusion proteins are well-established both for immunoprecipitation as well as subcellular localisation studies(248,249).

Affinity purification-mass spectrometry has been the gold standard of interaction proteomics for over a decade, replacing older yeast-2-hybrid methods(250). For this study, there was little value

in attempting to quantify interaction partners, as with an artificial overexpression system such as the pcDNA(+) plasmid used, quantification would be physiologically irrelevant. Therefore, it was decided to use the most sensitive mass spectrometer available, ThermoFisher's Orbitrap, available to identify proteins that interact with WISP-1(251).

It is clear from comparison of the identified interaction partners in HEK-293T cells to those in MRC-5 cells that the number of unique interaction partners is higher in the former. This may simply be because the overexpression plasmid was more efficiently transfected and expressed in HEK-293T cells, as evidenced by the higher fluorescence level observed in HEK-293T cells. More strikingly, however, there is little overlap between the protein interaction partners identified in HEK-293T cells and those identified in MRC-5 cells. This suggests that WISP-1 may play different roles in different cell types, and that its functional role is likely to be highly dependent on the protein complement of the cell it is expressed in. Its protein interaction network in IPF is likely to be very different for its network in cancer, for example.

The identity of the interaction partners is also of interest. The interaction partners identified in HEK-293T cells included multiple proteins associated with protein localisation to telomeres, including telomere length regulation protein TEL2 homolog (TELO2), as well as several members of the chaperonin containing tcp1 (CCT) complex suggesting that WISP-1 may play a role in regulating the function of this complex, which has varied roles in regulating cellular proliferation by altering telomere protein trafficking, as well as changing the folding of multiple cell signalling proteins associated with increased cell survival and resistance to apoptosis(252,253). Given that HEK-293T cells, while not a cancer cell line, display many similar properties to cancer cells in their fast growth, amorphous shape, and immortality, this finding may be more relevant to the role of WISP-1 in cancer(254). However, given the potential role telomeres may play in IPF pathogenesis as evidenced by several IPF susceptibility genes being associated with telomere renewal, Telomerase Reverse Transcriptase (TERT) and Telomerase RNA Component (TERC), it may be important for IPF pathogenesis as well if this interaction is present there(255).

The more prosaic alternative explanation for this interaction being identified is that the CCT chaperonin complex simply acts to aid WISP-1 protein folding, and there is no functional connection between the two. The large amount of WISP-1-GFP protein produced in HEK293T cells could lead to increased interaction with chaperones as cells respond to such a large amount of protein production. Additional follow-up work would be necessary to determine if this interaction does actually have consequences for WISP-1's functional role.

The protein interaction partners that MRC-5 and HEK-293T cells have in common also shed light on the functionality of WISP-1. Here, dermicidin is unusual in that it is an antimicrobial peptide, with little known function outside of its innate immune, bactericidal role. This potentially points to a role for WISP-1 within immune sensing. However, dermicidin has also been implicated in cell survival in cancer cells – 10% of breast cancer tumours expressed dermicidin and its expression was associated with cell growth and survival. In addition, it has been connected to neural cell survival(256–258). In prostate cancer, upregulation of DCD expression is associated with increased cell survival when cells are subjected to oxidative stress and hypoxia(259). However, dermicidin is a commonly identified contaminant in affinity purification-mass spectrometry results, so its presence is not necessarily indicative of it being a physiologically important interaction partner(260).

The other two WISP-1 protein interaction partners that are common between all MRC-5 samples and HEK-293T cells are prohibitin and VDAC1. These both have potentially important implications for WISP-1 functionality in IPF. Prohibitin is a mitochondrial inner membrane protein that acts as a scaffold for mitochondrial structure and may act as a molecular chaperone for other mitochondrial membrane proteins, including electron transport chain components(261,262). Members of the prohibitin family have diverse roles in a number of different processes, including cell cycle regulation, transcriptional regulation, cell signalling and apoptosis(239). The interaction of WISP-1 with prohibitin is thus likely of functional significance, but may be related to one or more of many different cellular processes. Prohibitin could not be identified in the GFP-tagged WISP-1 elution fractions when they were run on a western blot, however, so it may be a relatively low-affinity interaction partner whose level was sufficiently low not to show up on a western blot, but which could be detected using mass spectrometry. Prohibitin could readily be detected in input and flow-through fractions in MRC-5 lysates, suggesting that its expression is abundant in Further work is required to confirm or exclude this interaction partner.

VDAC1 is another mitochondrial protein with potentially important functional relevance for IPF. The WISP-1-VDAC1 interaction was also able to be verified by western blot, unlike prohibitin, strengthening the evidence that it is a genuine interaction partner for WISP-1. VDAC1 is a mitochondrial porin with many different potential substrates, including NADH, ATP/ADP, citrate, succinate, glutamate, pyruvate, and glucose, in addition to Ca2+, Mg2+, Cl−, K+, and Na+ ions(241,263). It can regulate several different mitochondrial and cellular processes, including apoptosis, and it interacts with hexokinase, the first enzyme in the glycolytic pathway. This interaction is associated with changes in glycolysis, oxidative phosphorylation, and cytochrome c release – the first step in apoptosis(264). The effect of VDAC1 on regulating glycolysis may have

relevance to myofibroblast activity, as inhibition of glycolytic activity reduces the contractile phenotype of myofibroblasts, as well as downregulating myofibroblast markers such as α-smooth muscle actin(265).

VDAC1 is hypoxia regulated, with a truncated form of the protein lacking a C-terminus being upregulated by HIF signalling. Given the previously demonstrated driving of WISP-1 via HIF signalling, as well as the importance of hypoxia in driving collagen crosslinking in IPF, it is likely VDAC1 may be important for fibroblast survival in IPF(266).

Basic fibroblast growth factor, or FGF2, is the final interaction partner that was investigated. It was also identified on a western blot of the MRC-5 elution fractions, suggesting that it is a genuine interaction partner for WISP-1. This interaction was also confirmed using a protein binding assay with recombinant protein shown in Figure 5.15. This assay was conducted using low-MW FGF2. However, there was an observed enrichment in high-MW FGF2 in elution fractions versus input and flow-through fractions, suggesting that high-MW FGF2 has greater affinity for WISP-1. This assay should be repeated with high-MW-FGF2 to identify if its interaction affinity with WISP-1 is actually greater. Unlike the other interaction partners identified, the interaction of FGF2 with a CCN protein family member, CTGF, has previously been identified(183). This interaction was localised to a particular domain, the thrombospondin response element domain, in CTGF, which has very high sequence homology with the corresponding domain in WISP-1 (Figure 5.14). Furthermore, in preliminary experiments addition of recombinant CTGF caused a reduction in the WISP-1-FGF2 interaction, suggesting that the two proteins might interact via the same location on the FGF2 molecule. Further binding assays would confirm this. For example, by coating ELISA plates with a fixed concentration of FGF2 and varying the concentrations of WISP-1 and CTGF would allow a more comprehensive assessment of the effect WISP-1, FGF2 and CTGF have on each other's binding. Other techniques, such as surface plasmon resonance, could be used to further characterise this WISP-1-FGF2 interaction in terms of affinity (on rate, off rate, Kd) in comparison with CTGF.

The CTGF-FGF2 interaction was found to induce cellular proliferation and production of matrix metalloproteinase enzymes in chondrocytes(183). The authors hypothesised that this effect was mediated via an interaction with the CTGF-FGF2 complex and the FGF2 receptor 1 (FGFR1). It is reasonable to hypothesise that the WISP-1-FGF2 interaction may work in a similar manner to mediate cellular proliferation and extracellular matrix production in IPF. Furthermore, the antifibrotic drug nintedanib works by affecting FGF receptor activity(21). FGF receptors are receptor tyrosine kinases to which FGF2 binds in order to mediate its function(267). It is plausible

that WISP-1 interacting with FGF2 could also affect FGFR2 signalling, potentially altering nintedanib's efficacy as a therapeutic. Furthermore, the observed enrichment of high-MW FGF2 – known to localise to the nucleus(245), coupled with the observation that WISP-1 variants 1, 3 and 4 localised to the nucleus in HeLa cells, suggests an alternative pathway for the WISP-1-FGF2 interaction to directly influence gene expression in the nucleus.

This chapter has identified several WISP-1 interaction partners which have the potential to affect the functionality of WISP-1 in IPF. However, to fully investigate how WISP-1 functions, more investigation will be required, both to confirm these interaction partners, and conduct functional assays into how they affect WISP-1 signalling in the context of pulmonary fibrosis.

6 Final Discussion

6.1 Summary of findings

The work presented in this thesis details three key areas which give information about the role of WISP-1 in IPF

Firstly, WISP-1 expression is localised to IPF mesenchymal cells, in particular HAS1 high fibroblasts and a subpopulation of myofibroblasts, both of which strongly express a number of profibrotic genes. Gene signatures associated with these cell types were identified in bulk RNAseq data from laser capture microdissected fibroblastic foci, suggesting that these cells are found in fibroblastic foci. RNAScope *in-situ* hybridisation confirmed that WISP-1 mRNA is expressed in fibroblastic foci. This contradicts the previous research of Berschneider and Königshoff, which identified WISP1 as being expressed in alveolar type II epithelial cells(191,192). This suggests a role for WISP-1 in promoting the fibroblastic focus, the epicentre of fibrosis in IPF.

Secondly, WISP-1 expression is driven by hypoxia signalling, specifically hypoxia inducible factor (HIF) signalling. This is a signalling pathway which has previously been associated with IPF pathogenesis, as well as driving tissue fibrosis in several other diseases(110). Hypoxia has specifically been implicated in dysfunctional regulation of collagen crosslinking leading to a higher proportion of bone-type pyridinoline crosslinks, leading to the stiffened collagen characteristic of IPF pathogenesis(65,96,215). This finding links WISP-1 with a known pro-fibrotic signalling pathway, as well as identifying a novel regulator of WISP1 expression, other than Wnt signalling.

Thirdly, WISP-1 interacts with several proteins with possible roles in promoting fibrosis. A tagged, overexpression, affinity-purification mass spectrometry workflow was used to identify WISP-1 interaction partners in two different cell lines, HEK-293T human embryonic kidney cells, and MRC-5 lung fibroblasts. The range of WISP-1 interaction partners identified by this approach differed significantly between the two cell lines, although there were some common interaction partners identified. In HEK-293T cells, WISP-1 was identified as interacting with a molecular chaperone complex involved in telomere maintenance, suggesting possible roles for WISP-1 in regulating cellular survival, proliferation and senescence. In MRC-5 cells, a more appropriate model for IPF given the localisation of WISP-1 to the fibroblastic foci explored in previous chapters, WISP-1 interacted with basic fibroblast growth factor (FGF2) a protein associated with fibroblast proliferation, as well as VDAC1 and prohibitin, two mitochondrial proteins which have previously been strongly associated with cellular survival and proliferation(239,268). The latter two interaction partners were also identified in HEK-293T cells. The interaction of WISP-1 with FGF2

and VDAC1 were confirmed by western blotting, and a binding curve for the FGF2-WISP-1 interaction was obtained. Notably, connective tissue growth factor, a CCN protein family member and known FGF2 interaction partner with strong sequence homology to WISP-1 partially interfered with this interaction, suggesting a similar binding mode between the two proteins. This is functionally is relevant to WISP-1 as the FGF2-CTGF interaction contributed to MMP9 production and proliferation in cultured chondrocytes(183). WISP-1 interacts with known proliferation and cell survival factors in cultured lung fibroblasts, and interacts with FGF2, a pro-fibrotic protein which is known to interact with a WISP-1 homologue.

In summary, WISP-1 is a protein which is expressed in pro-fibrotic fibroblasts which localise to fibroblastic foci, the hub of tissue fibrosis. This expression is driven by hypoxia, a known pro-fibrotic process. WISP-1 interacts with known proliferative and cell survival factors, as well as a fibroblast growth factor that is known to interact with a WISP-1 homologue and promotes proliferation and ECM production in chondrocytes.

Figure 6.1 shows a schematic of the mesenchymal cell development proposed in this discussion, and where WISP-1 may be acting to influence these phenotypes.

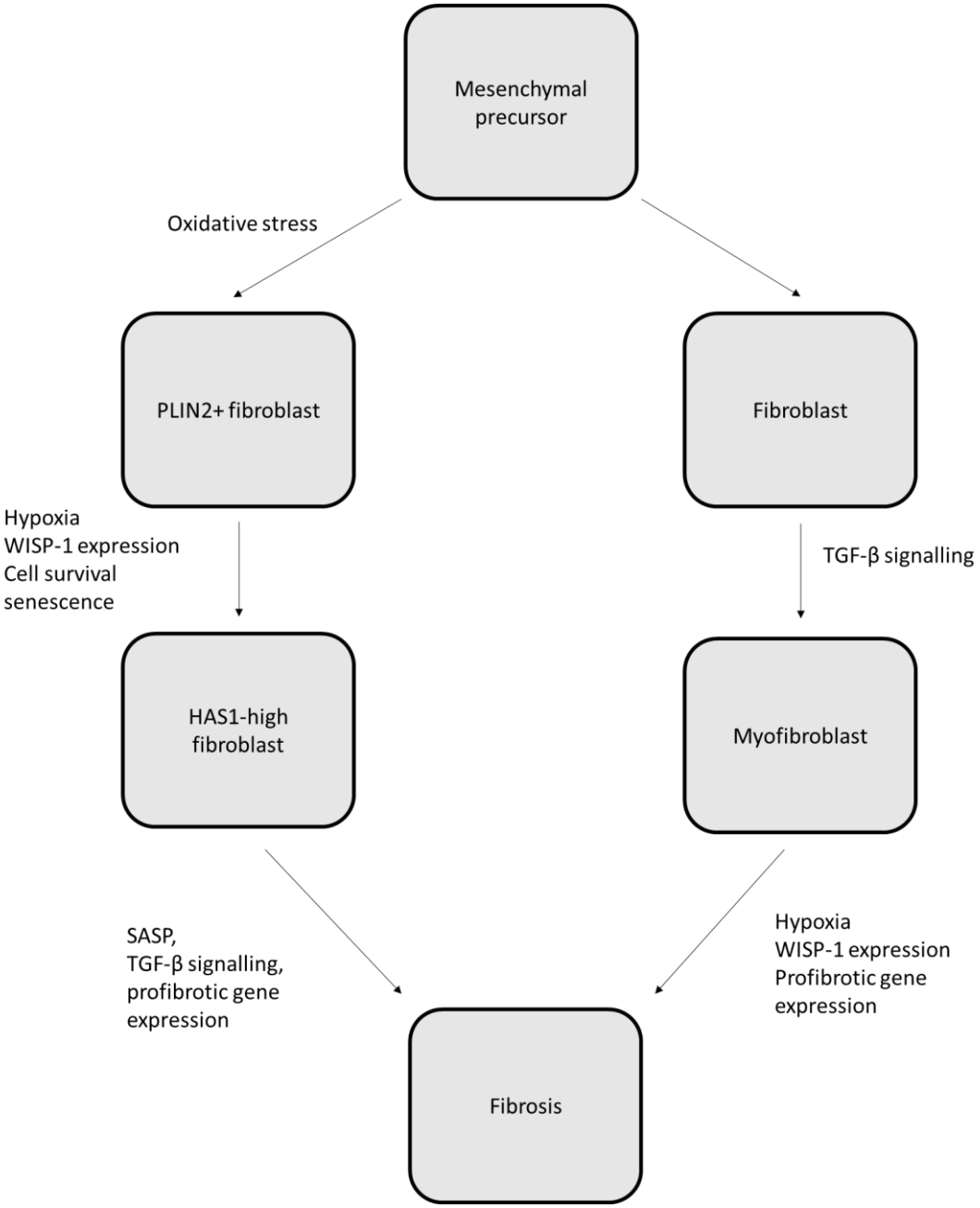


Figure 6.1: Schematic of proposed scheme for mesenchymal cell development in IPF, and where WISP-1 may act to promote fibrosis.

6.2 Limitations of the approaches used in this thesis

There are several limitations with the results presented here, both due to technical constraints as well as methodological choices.

Firstly, the lack of a control group for the fibroblastic focus laser capture microdissection data. A large part of the data presented in chapter 3 relies on patterns of gene expression in the laser capture microdissection fibroblastic focus samples. However, due to the nature of this sample, a comparator dataset from control lung mesenchyme is not available. It is unfeasible to extract fibroblasts from control lung tissue as they are embedded in the lung interstitium – any aggregation of fibroblasts into a focus that can be subjected to microdissection is necessarily pathological. This was part of the reasoning to turn to the single-cell RNAseq data – it allows calculation of gene expression signatures associated with difference fibroblast types which can then be mapped back onto the LCMD data, while also allowing direct comparison of different fibroblast populations. However, a differential gene expression analysis of the LCMD RNAseq data comparing control to IPF fibroblasts could not be performed.

The single cell data itself provided some methodological challenges. The nature of single-cell RNAseq data is a dataset with extremely high dimensionality (upwards of 100,000 cells/experiment), but with quite sparse data density – most cells return a counts value of 0 for most genes. The resulting relatively sparse counts matrix (in comparison to bulk RNAseq data) means that low-expressing genes may be missed entirely in many cells, even if in aggregate they would display expression of that gene. This is partially a consequence of the limits of sensitivity of single-cell RNA-seq – if only a small number of transcripts are present in a cell, they may not be sequenced. It is difficult to parse this data loss from the inherently stochastic nature of gene expression – certain cells may simply not be expressing a particular gene when fixed(269). Both these effects are magnified for low-expressing genes. The Seurat methodology used to analyse this data mitigates this to some extent, as does the very large dimensionality of the dataset – a “refuge in numbers” approach, but it is a significant downside of single cell RNAseq data compared to bulk(172).

6.3 Fibroblastic foci are complex, containing multiple cell types. WISP-1 is expressed in two very different fibroblast types

This research has added to our understanding of fibroblastic foci in IPF, characterising a number of different mesenchymal cell types from single cell RNAseq data, and then mapping the gene signatures associated with these cells back onto the bulk fibroblastic foci laser capture microdissection RNAseq data. WISP1 gene expression is associated with a subset of myofibroblasts which disproportionately express profibrotic genes, including several collagen genes, and genes associated with fibrosis such as SPARC, and the collagen crosslinking enzymes PLOD2 and LOXL2. WISP1 is also expressed in HAS-high fibroblasts, an IPF-derived population of mesenchymal cells which also produces a large number of extracellular matrix factors, as well as having a clear secretory phenotype, suggesting these are a subset of senescent fibroblasts. HAS1-high fibroblasts also express the oxidative stress responsive gene HMOX1.

WISP1 expression is thus associated with two different types of mesenchymal cell. These are very distinct in their gene expression signatures – they cluster far from each other on a t-SNE plot. However, both types of cell are associated both with an increased expression of profibrotic genes, and a high gene set variance analysis (GSVA) score for TGF-β-related genes.

The results presented in chapter 4 of this thesis, showing WISP-1 is induced by HIF stabilisation or hypoxia in cultured lung fibroblasts, as well as previous research demonstrating the importance of HIF pathway activation in fibrosis, suggests that these WISP-1-producing mesenchymal cells represent two different subpopulations of fibroblasts which are both responding to profibrotic TGF-β signalling and tissue HIF activation in the IPF lung.

This conclusion is supported by the presence of myofibroblasts which do not have the same profibrotic phenotype as the WISP-1 expressing myofibroblasts, and which have a commensurately lower HIF GSVA score. The single cell RNAseq data also presents a possible source cell type for the HAS1-high fibroblasts. There is a large population of PLIN2+ fibroblasts identified in this study, which cluster closely with the HAS1-high fibroblasts. These PLIN2+ fibroblasts have a similar gene expression signature to the HAS-high fibroblasts. However, there are also significant differences between the two. Notably, PLIN2+ fibroblasts express several genes associated with mitigating the negative effects of oxidative stress, including multiple metallothionein genes. Their characteristic expression of PLIN2 also suggests a connection with

lipofibroblasts, which help maintain epithelial barrier integrity in rodents. Although these cells lack some common features of lipofibroblasts – notably PPARγ expression, they may fulfil some of the same roles(153). The original paper which identified them used immunofluorescence staining to identify where these cell types are located in IPF lung tissue, and identified PLIN2+ fibroblasts in the interstitial space and surrounding alveoli(58). The cibersortX analysis of the LCMD RNAseq data did identify PLIN2+ fibroblast gene signatures in the fibroblastic focus data, as well as in the alveolar septae of both control and IPF lung tissue. Notably, the calculated proportion of PLIN2+ fibroblasts in this dataset was the only one out of all fibroblast populations which was highest in the control IPF septae samples, again suggesting that they represent a fibroblast type associated with alveolar maintenance.

This relationship between PLIN2 and HAS1-high fibroblasts is backed up by the pseudotime analysis of the single cell RNAseq data. This analysis shows clear similarity between the majority of PLIN2+ fibroblasts and HAS1-high fibroblasts, putting them in a separate cluster to the fibroblasts and myofibroblasts. The pseudotime score of these cells shows a clear progression from PLIN2+ to HAS1-high cells.

In summary, PLIN2+ fibroblasts have similar overall gene expression signatures to HAS1 high fibroblasts, but with a higher response to oxidative stress, they express PLIN2, known to be important to epithelial maintenance and present in lipofibroblasts in rodents, have a higher proportion in normal lung alveolar septae than IPF alveolar septae or fibroblastic foci, show progression from PLIN2+ to HAS high fibroblasts on a pseudotime analysis plot, and lack the secretory, senescent and profibrotic gene signatures present in HAS1-high fibroblasts. It is likely that HAS1-high fibroblasts are a profibrotic descendent of PLIN2+ fibroblasts.

6.4 The possible role of WISP-1 in different fibroblast subtypes

This study has identified WISP-1 as being a hypoxia-driven protein expressed in profibrotic mesenchymal cells. It is important to consider its functional role in these cells, by identifying the function of other CCN proteins,

CCN proteins mediate their diverse functions through interactions with other proteins. These interactions are mediated via the four domains that make up the functional unit of the CCN proteins. These are the insulin-like growth factor binding (IGFBP) domain, the Von Willebrand

factor type C (VWC) domain, the thrombospondin type-1 repeat (TSP) domain and the cysteine knot (CT) domain(121). This reliance on protein interactions to mediate their function is why a targeted protein interaction study was considered useful for investigating WISP-1 functionality.

Two functions that are common to multiple CCN proteins, including WISP-1, are regulation of cellular survival and regulation of bone development(246). These may both have relevance to how WISP-1 functions in IPF.

The importance of pathways which regulate bone development in IPF is of interest. Recent studies looking at collagen crosslinking in IPF have highlighted the importance of bone type pyridinoline crosslinks in forming the stiffened collagen fibrils which contribute to IPF pathogenesis(65). Gene ontology signatures associated with bone development and chondrogenesis were also found to be upregulated in bulk IPF transcriptomic data. Given the known role of WISP-1 in regulating bone development, as well as its expression in HAS1-high fibroblasts and profibrotic myofibroblasts, which both express elevated amounts of collagen and collagen crosslinking enzymes, there is potential for the known role of WISP-1 in bone development to be applicable to IPF pathogenesis as well. However, no calcification of tissue is seen in IPF, so while there may be production of bone-type collagen in fibrotic tissue, it is clear that mesenchymal cells in IPF do not fully differentiate to osteoblasts or chondrocytes.

The possible roles of WISP-1 which have been identified are backed up by the interaction partners identified in MRC-5 cells in chapter 4. Firstly, the interactions with VDAC1 and prohibitin are consistent with roles for WISP-1 in cellular survival. Both proteins have previously established roles in preventing apoptosis(241,261).

Prohibitin (PHB) is a protein inextricably linked with cellular proliferation. It is a mitochondrial membrane protein (much like VDAC1) which forms large, ring-shaped complexes in the mitochondrial membrane with the related protein PHB2. It is involved in mitochondrial structure formation, being important for cristae formation in mitochondrial development. It has also been identified as playing a role in the regulation of proliferation(239). Knockdown of PHB genes reduces cellular proliferation in mouse embryonic fibroblasts. Further studies were able to show that targeting of prohibitins to mitochondria is required for it to promote cellular proliferation(261). However, prohibitins have been identified as having potential anti-proliferative effects as well – they are negative regulators of E2 factor (E2F) mediated transcription – important for cell cycle regulation(270). There is therefore the possibility of the interaction of WISP-1 with prohibitin to induce proliferation by interacting with mitochondrial prohibitin complexes or downregulating the negative effect of prohibitin on E2F transcription. Without

further study, it is difficult to be more specific about a possible mechanism by which this interaction could influence cellular proliferation. The fact that PHB1 was one of the few common interaction partners between HEK-293T and MRC-5 cells suggests that it may be present in multiple different cellular contexts. However, prohibitin was not identified in a western blot analysis of GFP-trap co-IP elution fractions, despite being readily detectable in input and flow-through fractions, suggesting that its interaction with WISP-1 is lower affinity than other identified interaction partners.

VDAC1 knockdown in cancer cells leads to a decrease in their ability to proliferate and migrate, an effect attributed to a decrease in the availability of substrates for mitochondrial metabolism due to lack of VDAC1 porin activity(268). It is conceivable that WISP-1 may affect this process, binding to VDAC1 and altering its membrane transport efficacy. However, there is also the possibility that WISP-1 is important for the other process VDAC1 is strongly associated with – prevention of apoptosis(240,241). The mitochondrion is crucial for instigating intrinsic apoptosis; briefly, the B cell lymphoma-2 (Bcl-2) family of proteins, which contain both pro- and antiapoptotic members, associate with the mitochondrial outer membrane. Apoptotic signalling is initiated by the interaction of the Bcl-2 proteins BAX and BAC with VDAC1 leading to the depolarisation of the mitochondrial membrane. This triggers cytochrome C release, which initiates apoptosis. Bcl-2, an antiapoptotic Bcl-2 family member, also interacts with mitochondrial VDACs including VDAC1, and inhibits the action of BAX and BAC. The interaction of VDAC1 with these mitochondrial proteins is mediated via the N-terminal region of VDAC1(76,271).

Given the role of VDAC1 in regulating apoptosis, as a binding partner for both pro- and anti-apoptotic factors, and its central role in initiating apoptosis through mitochondrial membrane depolarisation, it is possible to envisage several mechanisms by which WISP-1 interaction affects the role of VDAC1 in regulating apoptosis. Firstly, WISP-1 could block the ion channel activity of VDAC1 directly, preventing mitochondrial membrane depolarisation and cytochrome C release. Secondly, WISP-1 could strengthen the VDAC1-Bcl-2 interaction, or weaken the VDAC1-BAX/BAC interaction, reducing the ability of apoptotic initiators to mediate cell death. Further study would be needed to investigate this, particularly investigation of which region of VDAC1 WISP-1 interacts. If WISP-1 interacts with the N-terminal region of VDAC1 known to be where the Bcl-2 proteins binds, it would add credence to the idea of WISP-1 mediating anti-apoptotic effects via altering the interaction of VDAC1 with Bcl-2 proteins(240).

This anti-apoptotic role of WISP-1 also ties into the interaction of WISP-1 with the chaperonin containing tcp1 (CCT) complex in HEK293T cells. This is a chaperone complex which is important

for regulating telomere assembly, as it aids the folding of TCAB1, a critical component of the telomerase enzyme complex which drives the recruitment of telomerase to cajal bodies, an important step in driving telomerase assembly(252,253). If this interaction does have functional significance, then WISP-1 may reduce the catalytic activity of telomerase by preventing folding of TCAB1, preventing telomere elongation and promoting senescence. This is inherently speculative, and the WISP-1-CCT interaction was not observed in MRC-5 cells so may have less relevance to lung fibroblasts, but it does tie into studies which have identified the genomic loci corresponding to the TERT and TERC genes, which code for other telomerase subunits, as being associated with increased incidence of IPF in several genome-wide association study(272).

This latter role – prevention of apoptosis – is important because it is relevant to the HAS1-high fibroblasts. As mentioned above, these have a highly secretory phenotype, which may be associated with senescence(202,203,243). An increase in WISP-1 levels due to HIF signalling in PLIN2+ fibroblasts, already subjected to oxidative stress, could lead to inhibition of apoptosis, leading to the PLIN2+ fibroblasts losing their protective, antifibrotic phenotype and becoming senescent, secretory and profibrotic under the influence of HIF and TGF-β signalling.

The interaction of WISP-1 with FGF2 has several potential functional consequences for WISP-1 expressing myofibroblasts or HAS1-high cells. In chapter 4, the similarities between this interaction and the interaction of FGF2 with CTGF were investigated. It is likely that these two interactions are mediated via the same protein domain (identified in CTGF as the TSP1 domain) as CTGF reduced WISP-1 binding to FGF2. In chondrocytes, the interaction of FGF2 and CTGF led to an increase in cellular proliferation, as well as an increase in matrix metalloproteinase 9 (MMP9) production(183). MMPs have been implicated in IPF pathogenesis. MMP9 has been linked with facilitating leukocyte migration by cleaving ECM proteins, cleaving and activating latent TGF-β and thus increasing TGF-β signalling(76,273). This could partially explain the strong increase in the TGF-β signalling GSVA score seen in the WISP1 expressing mesenchymal subpopulations in the single cell RNAseq data.

As the FGF2-CTGF interaction was associated with increased proliferation in chondrocytes, cartilage-producing cells, this may indicate that the WISP-1-FGF2 interaction in IPF lung fibroblasts is partially responsible for the adoption of the bone-type phenotype seen in the IPF lung. WISP-1 is a known driver of bone development, driving osteoblast differentiation and chondrocyte proliferation. The WISP-1-FGF2 interaction may be driving fibroblasts to proliferate in a similar manner to chondrocytes, and also driving the expression of genes associated with bone development, such as SPARC/osteonectin, and cartilage, such as asporin(188,274–276).

Beyond these specific effects observed from the CTGF-FGF2 interaction, it is important to consider the broader roles of FGF2, and how they might relate to IPF pathogenesis, as these are all processes that WISP-1 could affect. FGF2, like other fibroblast growth factors, is important in a variety of processes, including cell survival, division, differentiation and migration(277). It is associated with increased cellular proliferation in multiple cell types, including fibroblasts. The FGFR receptors are also targets for the antifibrotic receptor tyrosine kinase nintedanib(21). However, interestingly, several studies have identified FGF2 as having *antifibrotic* effects in skin fibrosis(278,279). Skin fibrosis shares several defining features with IPF, and certain forms of skin fibrosis, including keloid scarring, are associated with an overactive wound-healing response, in a similar manner to how IPF is hypothesised to begin(278). This effect was observed to be mediated by mitogen activated protein kinase (MAPK) signalling and reduced the effect of TGF-β signalling on skin fibroblasts(280). FGF2 also improves wound healing in burn victims, leading to lower scarring. This effect was also observed in hypertrophic (keloid) scarring, where FGF2 signalling was shown to induce hepatocyte growth factor (HGF) and MMP1 production. This was associated with a reduction in scar tissue formation suggesting a possible antifibrotic role for FGF2(278).

The suppression of myofibroblast differentiation associated with FGF2 may explain the distribution of FGF2 expression on a t-SNE plot of FGF2 gene expression in the single-cell RNA-seq data. FGF2 expression is virtually absent in myofibroblasts, but it is expressed in PLIN2+ and HAS1-high fibroblasts. If FGF2 signalling supresses the TGF-β driven myofibroblast phenotype, it is likely that myofibroblasts in IPF will express lower amounts of FGF2. The expression in HAS1-high fibroblasts suggests that these may be where the interaction with WISP-1 is taking place, although as both proteins can be extracellular, the interaction may not be confined to these cells. Pseudotime analysis suggests that myofibroblasts differentiate from more normal fibroblasts. Proliferation of these precursors in response to FGF2 signalling may increase the number of normal fibroblasts which can then differentiate into myofibroblasts in response to other signalling pathways such as TGF-β and HIF.

FGF2 notably has several different isoforms, and there was an apparent enrichment for high molecular weight FGF2 in co-IP elution fractions, compared to the ratio of low-to-high MW FGF2 in the co-IP input fraction. This has potential functional consequences for where this interaction occurs, and what functions WISP-1 may affect. High-MW FGF2 localises to the nucleus of cells, where it is a potent mitogen, leading to increased proliferation in several different cell types, including glioblastoma cells and fibroblasts(281–283). Nuclear localisation of several GFP-tagged WISP-1 isoforms, including full-length, was also observed in confocal microscopy of WISP1-GFP transfected HeLa cells. If WISP-1 preferentially interacts with high-MW FGF2, it may facilitate this

proliferation. It may be that high molecular weight FGF2 induces increased mesenchymal cell proliferation in fibrosis, aided by interacting with nuclear WISP-1. Other signalling pathways then may lead to differentiation of these cells into profibrotic mesenchymal cell types such as myofibroblasts and HAS1-high fibroblasts.

To summarise, the identified protein interaction partners of WISP-1 may have diverse functional consequences for the different types of IPF fibroblasts that have been identified from the single-cell RNAseq data. WISP-1 interacts with pro-proliferative and survival factors, possibly preventing protective PLIN2+ fibroblasts from undergoing apoptosis, leading to them to become senescent, profibrotic HAS1-high fibroblasts. WISP-1 also interacts with FGF2, a mitogen, which induces expression of the profibrotic metalloproteinase MMP9 via a similar interaction with CTGF, but also displays antifibrotic and anti-TGF-β signalling activity in skin fibrosis. WISP-1 preferentially interacts with a nuclear-localised, mitogenic, high molecular-weight isoform of FGF2. Based on these interaction partners, WISP-1 likely plays a diverse role in regulating cell survival and proliferation in mesenchymal cell populations in IPF.

6.5 The role of HIF signalling in driving pro-fibrotic phenotypes, and WISP-1 signalling in the context of HIF activation

The data presented in chapter 4 show that the primary driver of WISP-1 in primary IPF fibroblasts is HIF signalling. HIF signalling also drives the expression of the collagen crosslinking enzymes LOXL2 and PLOD2, whose expression correlates well with WISP-1 expression in IPF. It was also observed that primary IPF fibroblasts produce more WISP-1 at baseline under normoxic conditions, and show greater WISP-1 induction in response to activation of HIF signalling and exposure to hypoxia than their healthy counterparts. This suggests that IPF fibroblasts are potentiated to respond to HIF signalling more than healthy fibroblasts. The mechanism of this potentiation is unclear, but there are some potential candidates based on the data shown in this thesis.

Firstly, profibrotic signalling from TGF-β is well known to be upregulated in IPF, and known to drive fibrosis and the promotion of the myofibroblast phenotype – an assertion backed up by the strong colocalization of cells with high TGF-β GSVA scores and profibrotic phenotypes (and WISP1 expression) in the single cell RNAseq data(284). However, while there is a relationship between

TGF-β signalling and hypoxia signalling in the literature, it is more commonly the inverse of that suggested here; hypoxia signalling induces TGF-β signalling. A study conducted in Lewis lung carcinoma cells identified long-term exposure to hypoxia conditions upregulated TGF-β induced SMAD signalling, and a second study conducted in skin fibroblasts identified HIF1α signalling as inducing TGF-β signalling(285,286). This suggests that ongoing hypoxia signalling may contribute to TGF-β-induced fibrosis in IPF by enhancing TGF-β signalling. However, this also suggests that ongoing TGF-β signalling is unlikely to be the mechanism by which IPF fibroblasts are potentiated for hypoxia signalling.

Another possible mechanism for the observed potentiation is signalling from oxidative stress. Oxidative stress signalling has a well-established mechanism by which it can influence HIF signalling, via factor inhibiting HIF (FIH) signalling. FIH is an intracellular peroxide sensing molecule which responds to oxidative stress. It is a hydroxylase which negatively regulates HIF1α transcriptional activity(232). The presence of peroxide (a common product of oxidative stress, formed from combination of two •OH radicals), inhibits FIH activity, leading to reduced inhibition of HIF signalling, increasing the potential response to hypoxia, or leading to oxygen-independent HIF signalling(232). The large number of genes associated with oxidative stress sensing, especially metallothionein genes, present in PLIN2+ fibroblasts suggests that oxidative stress signalling may be important, at least in these cells. Oxidative stress has also been identified as being present in IPF tissue – a recent systematic review identified multiple oxidative stress biomarkers which are upregulated in IPF tissue, while previous research has demonstrated oxidative stress signalling to exacerbate different types of fibrosis, including pulmonary fibrosis(233,287). Oxidative stress signalling is a likely mechanism for the discrepancy between hypoxia-induced WISP-1 expression in IPF versus healthy fibroblasts.

A counterpoint to the above explanation is that PLOD2 expression in healthy lung fibroblasts showed a similar dose-response to IOX2 to IPF fibroblasts – suggesting that these cells can respond to hypoxia signalling, and that some other mechanism may be potentiating these cells to HIF-induced WISP1 expression specifically (a two-hit hypothesis, discussed below in regard to Wnt signalling). However, this could also be explained by an increased sensitivity of the PLOD gene to induction by HIF signalling.

To summarise, a probable mechanism for the increased WISP-1 expressed induced by HIF stabilisation in IPF fibroblasts compared to healthy fibroblasts is higher baseline HIF signalling in these fibroblasts. This may be induced by oxidative stress signalling, which is in evidence in single

cell RNAseq data for IPF fibroblasts and can increase HIF signalling through a known mechanism – the negative regulation of FIH.

6.6 Wnt signalling signatures are identifiable in IPF tissue, but WISP1 expression was not Wnt-inducible.

Another discussion point this research has identified is that WISP-1 is not induced by Wnt3a or 5a signalling in IPF fibroblasts, but by HIF stabilisation. This runs counter to previous research, which identified WISP-1 as a Wnt-inducible protein. Furthermore, although WISP-1 induction may be different in different cell types, there is a large body of research identifying active Wnt signalling as being important in IPF(131,132,146). LCMD RNAseq expression data demonstrated an upregulation in genes associated with Wnt protein binding in fibroblastic foci, and RNAscope *in-situ* hybridisation analysis of WNT5A expression localises to fibroblastic foci in IPF, suggesting that there is active Wnt signalling in fibroblastic foci. However, no Wnt signalling gene expression signatures were identified in the single cell data, and cultured fibroblasts treated with Wnt ligands did not show an increase in WISP-1 expression. The principal Wnt-related gene identified in this thesis was SFRP2, a Wnt antagonist which may also be induced by Wnt signalling, was upregulated in bulk RNAseq, microarray and LCMD fibroblast focus data. This may suggest that there is some Wnt signalling present in these datasets, but it is difficult to parse out using single cell RNAseq data. However, it is possible that WISP-1 expression requires more than one input stimulus, and that constitutively active Wnt signalling is required in addition to hypoxia signalling for WISP-1 production – two hits from different signalling pathways are required to induce WISP1 expression. This would also provide an alternative, two-hit explanation for the potentiation of IPF fibroblasts for WISP-1 production – they are not necessarily less sensitive to HIF signalling, but require activation of another signalling pathway before inducing WISP1 expression. Given Wnt signalling's position as the pre-eminent WISP-1 inducer, it is a good candidate for further investigation – transfection of a TOPflash luciferase Wnt signalling plasmid into IPF and normal fibroblasts would give an idea of the level of constitutive Wnt signalling activity in these cells.

6.7 Splice variants

This study also investigated WISP-1 splice variants, of which there are four which can be translated, including the full-length protein(156). The BaseScope images shown in chapter 4 demonstrate that mRNA corresponding to variant 2, at least, is present in IPF tissue. The chapter

4 results also demonstrated that hypoxia can induce expression of all four splice variants in IPF lung fibroblasts.

This may have significance for how WISP-1 functions in IPF. WISP-1 is a highly modular protein, and it is likely that different functions of WISP-1 are mediated by different WISP-1 domains; a common feature of CCN proteins. Therefore, splice variants which lack certain domains will also lack any functionality that these domains confer. Ambler (2016) identified that MRC-5 cells transfected with plasmid expressing full-length WISP-1 showed very strong induction of WISP1 variant 4, suggesting that this variant may be upregulated in IPF fibroblasts expressing WISP-1(143).

The domain specificity of the WISP-1 interaction partners identified in chapter 5 is unknown. However, the paper which identified the CTGF-FGF2 interaction localised it to the TSP1 domain of CTGF – a domain which WISP-1 shares and has significant sequence homology with. If WISP-1 does interact with FGF2 via this domain, then WISP-1 variant 2, which also contains this domain and is expressed in IPF tissue may also be able to mediate this interaction.

6.8 WISP-1 may function in an intracellular and a paracrine manner

The research presented here has largely focussed on the interaction partners of WISP-1 in the intracellular space – although some, such as FGF2, are also capable of being secreted and functioning extracellularly. However, most prior research on WISP-1 has focussed on its functional role outside the cell, as a matricellular protein. This thesis also contradicts previous research on the role of WISP-1 in IPF conducted by Berschneider et al. suggesting that WISP-1 is produced by alveolar type II cells, and its primary fibrogenic role is in driving epithelial to mesenchymal transition in these cells(192). Immunohistochemical staining conducted by them identified WISP-1 as localising to the fibrotic epithelium of IPF tissue, again suggesting a role for WISP-1 in signalling to ATII cells in fibrosis.

It is possible that these findings could be reconciled by WISP-1 acting in a paracrine manner on ATII cells, but still being produced in fibroblasts within fibroblastic foci. Localisation data used for this study (RNAscope, LCMD RNAseq, single cell RNAseq data), focussed on WISP1 mRNA expression, not protein localisation. It is conceivable that WISP-1 protein localises to the ATII cells overlying the fibroblastic foci and contributes to the alveolar destruction observed in these cells. CCN proteins are known for acting in a paracrine manner, so this behaviour would have precedent(246).

However, this study has identified several intracellular interaction partners in lung fibroblasts which likely have functional consequences for fibroblasts in IPF, and even if WISP-1 does signal in a paracrine manner, that signalling will likely affect fibroblasts as well as they are the cells which actually produce WISP-1 in IPF. Based on the studies in this thesis, it is proposed that the primary role of WISP-1 in IPF is in mesenchymal cell populations, and that any effect on ATII cells is likely to be secondary.

6.9 Future Work

This research has provided a number of novel findings concerning WISP1 expression and function leading to new questions as to how WISP-1 functions in IPF. The following consists of a list of these, and possible follow-up work which could be done to answer them.

Firstly, it is unclear what the downstream effect of WISP-1 signalling is in lung fibroblasts – how does it affect gene expression, and what is the phenotype of cells affected by it. Secondly, it is unclear what the transcriptional differences between cultured healthy and IPF lung fibroblasts are – what is the mechanism by which IPF fibroblasts are potentiated to be more prone to HIF-induced WISP-1 induction? An RNAseq study of cultured IPF and healthy fibroblasts, with and without HIF stimulus via IOX2, and with and without WISP-1 knockdown, is currently underway, with the goal of answering these questions. This will allow identification of how the phenotype of these fibroblasts changes in the absence of WISP-1 signalling, and what effect this has on the expression of profibrotic genes. This study will address the question of where WISP-1 sits in the context of profibrotic signalling pathways – which profibrotic genes are driven by WISP-1 signalling. This functional transcriptomic study will complement the more mechanistic protein interaction partner study presented in this thesis.

Another outstanding question is whether WISP-1 acts in a paracrine manner on either fibroblasts or ATII cells in IPF. Immunohistochemistry studies to identify where WISP-1 localises to in IPF tissue, coupled with functional studies looking at EMT markers in ATII cells treated with WISP-1 would allow this to be investigated.

Additionally, the protein interaction partner study was conducted on MRC-5 fibroblasts – a lung fibroblast cell line, but at one remove from primary IPF fibroblasts. Conducting a similar affinity purification-mass spectrometry study on primary IPF fibroblasts after HIF activation to induce increased WISP1 expression would be a way both to validate the interaction partners identified here, as well as to potentially identify new ones which are IPF specific.

In a similar vein, this work was conducted using 2D cell cultures. These have some limitations, especially for fibroblasts, as mechanical stretch – a function of growth substrate stiffness – is a well-attested mechanism of signalling to fibroblasts, and cell culture plastic is far stiffer than even highly fibrotic IPF lung ECM. The University of Southampton has developed a 3D cell culture model for lung fibroblasts which closely resembles a fibroblast focus, producing highly stiffened ECM. Functional investigation of WISP-1 signalling in this model would better approximate the physiological environment of a fibroblastic foci.

There are numerous other lines of enquiry which could be pursued, such as investigation of synergistic effects of profibrotic signalling pathways such as oxidative stress, hypoxia and TGF-β signalling in IPF fibroblasts to investigate the potentiation of IPF fibroblasts to WISP1 expression, or investigation of the crosstalk between WISP-1 producing fibroblasts and ATII cells by growing ATII cells in conditioned media from WISP-1 producing fibroblasts.

In conclusion, the work presented in this thesis has substantially expanded our knowledge of the role of WISP-1 in IPF, having identified its localisation, drivers and functional interaction partners. This research will provide a base for future research into WISP-1, to further identify how it contributes to IPF pathogenesis.

7 Bibliography

1. Castranova V, Rabovsky J, Tucker JH, Miles PR. The alveolar type II epithelial cell: A multifunctional pneumocyte. Toxicology and Applied Pharmacology. 1988 May 1;93(3):472–83.

2. Churg A. The uptake of mineral particles by pulmonary epithelial cells. Am J Respir Crit Care Med. 1996 Oct;154(4 Pt 1):1124–40.

3. Barkauskas CE, Cronce MJ, Rackley CR, Bowie EJ, Keene DR, Stripp BR, et al. Type 2 alveolar cells are stem cells in adult lung. J Clin Invest. 2013 Jul 1;123(7):3025–36.

4. Varon J, Marik PE, Bisbal ZD. CHAPTER 1 - Restrictive Diseases. In: Papadakos PJ, Lachmann B, Visser-Isles L, editors. Mechanical Ventilation. Philadelphia: W.B. Saunders; 2008. p. 3–10.

5. Raghu G, Weycker D, Edelsberg J, Bradford WZ, Oster G. Incidence and Prevalence of Idiopathic Pulmonary Fibrosis. American Journal of Respiratory and Critical Care Medicine. 2006 Oct;174(7):810–6.

6. Wolters PJ, Collard HR, Jones KD. Pathogenesis of Idiopathic Pulmonary Fibrosis. Annual review of pathology. 2014 Sep;9:157–79.

7. Chung JH, Goldin JG. Interpretation of HRCT Scans in the Diagnosis of IPF: Improving Communication Between Pulmonologists and Radiologists. Lung. 2018;196(5):561–7.

8. Wuyts WA, Cavazza A, Rossi G, Bonella F, Sverzellati N, Spagnolo P. Differential diagnosis of usual interstitial pneumonia: when is it truly idiopathic? European Respiratory Review. 2014 Sep 1;23(133):308–19.

9. Lynch DA, Sverzellati N, Travis WD, Brown KK, Colby TV, Galvin JR, et al. Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. Lancet Respir Med. 2018 Feb;6(2):138–53.

10. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. The Lancet. 2017;389(10082):1941–52.

11. Usual interstitial pneumonia. In: Wikipedia . 2020 [cited 2020 Dec 21]. Available from: https://en.wikipedia.org/w/index.php?title=Usual_interstitial_pneumonia&oldid=992869778

12. Luppi F, Cerri S, Taddei S, Ferrara G, Cottin V. Acute exacerbation of idiopathic pulmonary fibrosis: a clinical review. Internal and Emergency Medicine. 2015;10(4):401–11.

13. Baumgartner KB, Samet JM, Stidley CA, Colby TV, Waldron JA. Cigarette smoking: a risk factor for idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 1997 Jan 1;155(1):242–8.

14. Taskar VS, Coultas DB. Is Idiopathic Pulmonary Fibrosis an Environmental Disease? Proc Am Thorac Soc. 2006 Jun 1;3(4):293–8.

15. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet. 2013 Jun;45(6):613–20.

16. Roy MG, Livraghi-Butrico A, Fletcher AA, McElwee MM, Evans SE, Boerner RM, et al. Muc5b Is Required for Airway Defense. Nature. 2014 Jan 16;505(7483):412–6.

17. DSP - Desmoplakin - Homo sapiens (Human) - DSP gene & protein . [cited 2018 Nov 30]. Available from: https://www.uniprot.org/uniprot/P15924

18. Hughes G, Toellner H, Morris H, Leonard C, Chaudhuri N. Real World Experiences: Pirfenidone and Nintedanib are Effective and Well Tolerated Treatments for Idiopathic Pulmonary Fibrosis. Barnes D, editor. Journal of Clinical Medicine. 2016 Sep;5(9):78.

19. Sario AD, Bendia E, Baroni GS, Ridolfi F, Casini A, Ceni E, et al. Effect of pirfenidone on rat hepatic stellate cell proliferation and collagen production. Journal of Hepatology. 2002 Nov 1;37(5):584–91.

20. Lin X, Yu M, Wu K, Yuan H, Zhong H. Effects of Pirfenidone on Proliferation, Migration, and Collagen Contraction of Human Tenon's Fibroblasts In Vitro. Invest Ophthalmol Vis Sci. 2009 Aug 1;50(8):3763–70.

21. Wollin L, Maillet I, Quesniaux V, Holweg A, Ryffel B. Antifibrotic and anti-inflammatory activity of the tyrosine kinase inhibitor nintedanib in experimental models of lung fibrosis. J Pharmacol Exp Ther. 2014 May;349(2):209–20.

22. Gurujeyalakshmi G, Hollinger MA, Giri SN. Pirfenidone inhibits PDGF isoforms in bleomycin hamster model of lung fibrosis at the translational level. Am J Physiol. 1999 Feb;276(2):L311-318.

23. Schaefer CJ, Ruhrmund DW, Pan L, Seiwert SD, Kossen K. Antifibrotic activities of pirfenidone in animal models. European Respiratory Review. 2011 Jun 1;20(120):85–97.

24. Hall CL, Wells AR, Leung KP. Pirfenidone reduces profibrotic responses in human dermal myofibroblasts, in vitro. Laboratory Investigation. 2018 May;98(5):640–55.

25. Izbicki G, Segel M, Christensen T, Conner M, Breuer R. Time course of bleomycin-induced lung fibrosis. Int J Exp Pathol. 2002 Jun;83(3):111–9.

26. Khalil N, O'Connor R. The Role of TGF-β in Bleomycin Induced Pulmonary Fibrosis. In: Jakowlew SB, editor. Transforming Growth Factor-β in Cancer Therapy, Volume I: Basic and Clinical Biology . Totowa, NJ: Humana Press; 2008. p. 581–94. (Cancer Drug Discovery and Development).

27. Moeller A, Ask K, Warburton D, Gauldie J, Kolb M. The bleomycin animal model: a useful tool to investigate treatment options for idiopathic pulmonary fibrosis? Int J Biochem Cell Biol. 2008;40(3):362–82.

28. Tashiro J, Rubio GA, Limper AH, Williams K, Elliot SJ, Ninou I, et al. Exploring Animal Models That Resemble Idiopathic Pulmonary Fibrosis. Front Med (Lausanne) . 2017 Jul 28;4.

29. Barbarin V, Arras M, Misson P, Delos M, McGarry B, Phan SH, et al. Characterization of the effect of interleukin-10 on silica-induced lung fibrosis in mice. Am J Respir Cell Mol Biol. 2004 Jul;31(1):78–85.

30. Naik PN, Horowitz JC, Moore TA, Wilke CA, Toews GB, Moore BB. Pulmonary Fibrosis Induced by γ-Herpesvirus in Aged Mice Is Associated With Increased Fibroblast

Responsiveness to Transforming Growth Factor-β. J Gerontol A Biol Sci Med Sci. 2012 Jul;67(7):714–25.

31. Sime PJ, Xing Z, Graham FL, Csaky KG, Gauldie J. Adenovector-mediated gene transfer of active transforming growth factor-beta1 induces prolonged severe fibrosis in rat lung. J Clin Invest. 1997 Aug 15;100(4):768–76.

32. Chambers RC, Mercer PF. Mechanisms of Alveolar Epithelial Injury, Repair, and Fibrosis. Ann Am Thorac Soc. 2015 Mar;12(Suppl 1):S16–20.

33. Burgy O, Königshoff M. The WNT signaling pathways in wound healing and fibrosis. Matrix Biology. 2018 Aug 1;68–69:67–80.

34. Tatler AL, Jenkins G. TGF-β Activation and Lung Fibrosis. Proceedings of the American Thoracic Society. 2012 Jul;9(3):130–6.

35. Kasper M, Barth K. Potential contribution of alveolar epithelial type I cells to pulmonary fibrosis. Biosci Rep . 2017 Nov 21;37(6).

36. Li F, Xu X, Geng J, Wan X, Dai H. The autocrine CXCR4/CXCL12 axis contributes to lung fibrosis through modulation of lung fibroblast activity. Experimental and Therapeutic Medicine. 2020 Mar 1;19(3):1844–54.

37. Kim KK, Wei Y, Szekeres C, Kugler MC, Wolters PJ, Hill ML, et al. Epithelial cell α3β1 integrin links β-catenin and Smad signaling to promote myofibroblast formation and pulmonary fibrosis. J Clin Invest. 2009 Jan 5;119(1):213–24.

38. Munger JS, Huang X, Kawakatsu H, Griffiths MJ, Dalton SL, Wu J, et al. The integrin alpha v beta 6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis. Cell. 1999 Feb 5;96(3):319–28.

39. Sugahara K, Tokumine J, Teruya K, Oshiro T. Alveolar epithelial cells: differentiation and lung injury. Respirology. 2006 Jan;11 Suppl:S28-31.

40. Strieter RM. What differentiates normal lung repair and fibrosis? Inflammation, resolution of repair, and fibrosis. Proc Am Thorac Soc. 2008 Apr 15;5(3):305–10.

41. Chua F, Laurent GJ. FIBROBLASTS. In: Laurent GJ, Shapiro SD, editors. Encyclopedia of Respiratory Medicine. Oxford: Academic Press; 2006. p. 213–9.

42. Burgstaller G, Oehrle B, Gerckens M, White ES, Schiller HB, Eickelberg O. The instructive extracellular matrix of the lung: basic composition and alterations in chronic lung disease. European Respiratory Journal. 2017 Jul 1;50(1):1601805.

43. Jones MG, Fabre A, Schneider P, Cinetto F, Sgalla G, Mavrogordato M, et al. Three-dimensional characterization of fibroblast foci in idiopathic pulmonary fibrosis. JCI Insight. ;1(5). 2016 Apr 21. 1(5):e86375.

44. Xia H, Bodempudi V, Benyumov A, Hergert P, Tank D, Herrera J, et al. Identification of a Cell-of-Origin for Fibroblasts Comprising the Fibrotic Reticulum in Idiopathic Pulmonary Fibrosis. The American Journal of Pathology. 2014 May;184(5):1369–83.

45. Cool CD, Groshong SD, Rai PR, Henson PM, Stewart JS, Brown KK. Fibroblast Foci Are Not Discrete Sites of Lung Injury or Repair. Am J Respir Crit Care Med. 2006 Sep 15;174(6):654–8.

46. Andersson-Sjöland A, de Alba CG, Nihlberg K, Becerril C, Ramírez R, Pardo A, et al. Fibrocytes are a potential source of lung fibroblasts in idiopathic pulmonary fibrosis. The International Journal of Biochemistry & Cell Biology. 2008;40(10):2129–40.

47. Hinz B, Phan SH, Thannickal VJ, Galli A, Bochaton-Piallat M-L, Gabbiani G. The Myofibroblast: One Function, Multiple Origins. The American Journal of Pathology. 2007 Jun;170(6):1807–16.

48. Rock JR, Barkauskas CE, Cronce MJ, Xue Y, Harris JR, Liang J, et al. Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. Proc Natl Acad Sci U S A. 2011 Dec 27;108(52):E1475–83.

49. Hung C, Linn G, Chow Y-H, Kobayashi A, Mittelsteadt K, Altemeier WA, et al. Role of Lung Pericytes and Resident Fibroblasts in the Pathogenesis of Pulmonary Fibrosis. Am J Respir Crit Care Med. 2013 Oct 1;188(7):820–30.

50. Hinz B, Lagares D. Evasion of apoptosis by myofibroblasts: a hallmark of fibrotic diseases. Nature Reviews Rheumatology. 2020 Jan;16(1):11–31.

51. Moore MW, Herzog EL. Regulation and Relevance of Myofibroblast Responses in Idiopathic Pulmonary Fibrosis. Curr Pathobiol Rep. 2013 Sep;1(3):199–208.

52. Desmoulière A, Geinoz A, Gabbiani F, Gabbiani G. Transforming growth factor-beta 1 induces alpha-smooth muscle actin expression in granulation tissue myofibroblasts and in quiescent and growing cultured fibroblasts. J Cell Biol. 1993 Jul;122(1):103–11.

53. Froese AR, Shimbori C, Bellaye P-S, Inman M, Obex S, Fatima S, et al. Stretch-induced Activation of Transforming Growth Factor-β1 in Pulmonary Fibrosis. Am J Respir Crit Care Med. 2016 Jan 15;194(1):84–96.

54. McGowan SE, Torday JS. The Pulmonary Lipofibroblast (lipid Interstitial Cell) and Its Contributions to Alveolar Development. Annual Review of Physiology. 1997;59(1):43–62.

55. Vaccaro C, Brody JS. Ultrastructure of developing alveoli. I. The role of the interstitial fibroblast. Anat Rec. 1978 Dec;192(4):467–79.

56. Torday JS, Sun H, Qin J. Prostaglandin E2 integrates the effects of fluid distension and glucocorticoid on lung maturation. American Journal of Physiology-Lung Cellular and Molecular Physiology. 1998 Jan 1;274(1):L106–11.

57. Rehan VK, Sugano S, Wang Y, Santos J, Romero S, Dasgupta C, et al. Evidence for the Presence of Lipofibroblasts in Human Lung. Experimental Lung Research. 2006 Jan 1;32(8):379–93.

58. Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. Science Advances. 2020 Jul 1;6(28):eaba1972.

59. Suki B, Ito S, Stamenovic D, Lutchen KR, Ingenito EP. Biomechanics of the lung parenchyma: critical roles of collagen and mechanical forces. J Appl Physiol (1985). 2005 May;98(5):1892–9.

60. Estany S, Vicens V, Llatjós R, Penín R, Escobar I, Xaubet A, et al. Extracellular matrix profile of lung in idiopathic pulmonary fibrosis. European Respiratory Journal . 2011 Sep 1;38(Suppl 55).

61. Ricard-Blum S. The Collagen Family. Cold Spring Harb Perspect Biol . 2011 Jan;3(1).

62. Vallet SD, Ricard-Blum S. Lysyl oxidases: from enzyme activity to extracellular matrix cross-links. Essays Biochem. 2019 Sep 13;63(3):349–64.

63. Knott L, Bailey AJ. Collagen cross-links in mineralizing tissues: A review of their chemistry, function, and clinical relevance. Bone. 1998 Mar 1;22(3):181–7.

64. Robins SP. Biochemistry and functional significance of collagen cross-linking. Biochem Soc Trans. 2007 Nov;35(Pt 5):849–52.

65. Jones MG, Andriotis OG, Roberts JJ, Lunn K, Tear VJ, Cao L, et al. Nanoscale dysregulation of collagen structure-function disrupts mechano-homeostasis and mediates pulmonary fibrosis. Vunjak-Novakovic G, editor. eLife. 2018 Jul 3;7:e36354.

66. Faffe DS, Zin WA. Lung parenchymal mechanics in health and disease. Physiol Rev. 2009 Jul;89(3):759–75.

67. Wagenseil JE, Mecham RP. New insights into elastic fiber assembly. Birth Defects Res C Embryo Today. 2007 Dec;81(4):229–40.

68. Ludwig MS. Chapter 5 - Proteoglycans in the Lung. In: Garg HG, Cowman MK, Hales CA, editors. Carbohydrate Chemistry, Biology and Medical Applications . Oxford: Elsevier; 2008. p. 113–31.

69. Wells RG. Tissue mechanics and fibrosis. Biochim Biophys Acta. 2013 Jul;1832(7):884–90.

70. Hashimoto K, Otero M, Imagawa K, de Andrés MC, Coico JM, Roach HI, et al. Regulated transcription of human matrix metalloproteinase 13 (MMP13) and interleukin-1β (IL1B) genes in chondrocytes depends on methylation of specific proximal promoter CpG sites. J Biol Chem. 2013 Apr 5;288(14):10061–72.

71. Lindsey ML, Zouein FA, Tian Y, Iyer RP, Brás LE de C. Osteopontin is proteolytically processed by matrix metalloproteinase 9. Canadian journal of physiology and pharmacology. 2015 Oct;93(10):879.

72. Selman M, Pardo A, Barrera L, Estrada A, Watson SR, Wilson K, et al. Gene Expression Profiles Distinguish Idiopathic Pulmonary Fibrosis from Hypersensitivity Pneumonitis. American Journal of Respiratory and Critical Care Medicine. 2006 Jan;173(2):188–98.

73. Bauer Y, White ES, de Bernard S, Cornelisse P, Leconte I, Morganti A, et al. MMP-7 is a predictive biomarker of disease progression in patients with idiopathic pulmonary fibrosis. ERJ Open Res. 2017 Jan;3(1).

74. Zuo F, Kaminski N, Eugui E, Allard J, Yakhini Z, Ben-Dor A, et al. Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. Proc Natl Acad Sci U S A. 2002 Apr 30;99(9):6292–7.

75. Pardo A, Selman M, Kaminski N. Approaching the degradome in idiopathic pulmonary fibrosis. The International Journal of Biochemistry & Cell Biology. 2008 Jun 1;40(6):1141–55.

76. Pardo A, Cabrera S, Maldonado M, Selman M. Role of matrix metalloproteinases in the pathogenesis of idiopathic pulmonary fibrosis. Respir Res . 2016;17.

77. Guillotin D, Taylor AR, Platé M, Mercer PF, Edwards LM, Haggart R, et al. Transcriptome analysis of IPF fibroblastic foci identifies key pathways involved in fibrogenesis. bioRxiv. 2020 Mar 11;2020.03.10.984955.

78. Yamashita M, Yamauchi K, Chiba R, Iwama N, Date F, Shibata N, et al. The definition of fibrogenic processes in fibroblastic foci of idiopathic pulmonary fibrosis based on morphometric quantification of extracellular matrices. Hum Pathol. 2009 Sep;40(9):1278–87.

79. Katzenstein A-LA, Myers JL. Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 1998 Apr 1;157(4):1301–15.

80. Noguchi S, Saito A, Mikami Y, Urushiyama H, Horie M, Matsuzaki H, et al. TAZ contributes to pulmonary fibrosis by activating profibrotic functions of lung fibroblasts. Sci Rep. 2017 Feb 14;7:42595.

81. Tschumperlin DJ, Ligresti G, Hilscher MB, Shah VH. Mechanosensing and fibrosis. J Clin Invest. 128(1):74–84.

82. Scotton CJ, Chambers RC. Molecular Targets in Pulmonary Fibrosis. Chest. 2007;132(4):1311–21.

83. Annes JP, Munger JS, Rifkin DB. Making sense of latent TGFβ activation. Journal of Cell Science. 2003 Jan;116(2):217 LP – 224.

84. Massagué J. TGF-β Signal Transduction. Annual Review of Biochemistry. 1998 Jun;67(1):753–91.

85. Massagué J. TGFβ signalling in context. Nature Reviews Molecular Cell Biology. 2012 Oct;13(10):616–30.

86. Branton MH, Kopp JB. TGF-β and fibrosis. Microbes and Infection. 1999;1(15):1349–65.

87. Clarke DL, Carruthers AM, Mustelin T, Murray LA. Matrix regulation of idiopathic pulmonary fibrosis: the role of enzymes. Fibrogenesis & Tissue Repair. 2013 Nov 26;6(1):20.

88. Willis BC, Borok Z. TGF-β-induced EMT: mechanisms and implications for fibrotic lung disease. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2007 Sep 1;293(3):L525–34.

89. Gharaee-Kermani M, Hu B, Phan SH, Gyetko MR. Recent advances in molecular targets and treatment of idiopathic pulmonary fibrosis: focus on TGFbeta signaling and the myofibroblast. Curr Med Chem. 2009;16(11):1400–17.

90. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. The Journal of Clinical Investigation. 2009 Jun;119(6):1420–8.

91. Salton F, Volpe MC, Confalonieri M. Epithelial–Mesenchymal Transition in the Pathogenesis of Idiopathic Pulmonary Fibrosis. Medicina (Kaunas). 2019 Mar 28;55(4).

92. Kim KK, Kugler MC, Wolters PJ, Robillard L, Galvez MG, Brumwell AN, et al. Alveolar epithelial cell mesenchymal transition develops in vivo during pulmonary fibrosis and is regulated by the extracellular matrix. Proceedings of the National Academy of Sciences of the United States of America. 2006 Aug;103(35):13180–5.

93. Kalluri R, Neilson EG. Epithelial-mesenchymal transition and its implications for fibrosis. Journal of Clinical Investigation. 2003 Dec;112(12):1776–84.

94. Tzouvelekis A, Harokopos V, Paparountas T, Oikonomou N, Chatziioannou A, Vilaras G, et al. Comparative Expression Profiling in Pulmonary Fibrosis Suggests a Role of Hypoxia-inducible Factor-1α in Disease Pathogenesis. Am J Respir Crit Care Med. 2007 Dec 1;176(11):1108–19.

95. Aquino-Gálvez A, González-Ávila G, Jiménez-Sánchez LL, Maldonado-Martínez HA, Cisneros J, Toscano-Marquez F, et al. Dysregulated expression of hypoxia-inducible factors augments myofibroblasts differentiation in idiopathic pulmonary fibrosis. Respiratory Research. 2019 Jun 24;20(1):130.

96. Brereton CJ, Yao L, Bell J, Conforti F, Alzetani A, Fabre A, et al. Hypoxia-inducible factor pathway activation promotes bone-type collagen cross-linking in Idiopathic Pulmonary Fibrosis. European Respiratory Journal . 2019 Sep 28;54(suppl 63).

97. Robinson CM, Neary R, Levendale A, Watson CJ, Baugh JA. Hypoxia-induced DNA hypermethylation in human pulmonary fibroblasts is associated with Thy-1 promoter methylation and the development of a pro-fibrotic phenotype. Respir Res. 2012 Aug 31;13:74.

98. Berg JM, Tymoczko JL, Stryer L, Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th ed. W H Freeman; 2002.

99. Lee JW, Ko J, Ju C, Eltzschig HK. Hypoxia signaling in human diseases and therapeutic targets. Experimental & Molecular Medicine. 2019 Jun;51(6):1–13.

100. Fong G-H, Takeda K. Role and regulation of prolyl hydroxylase domain proteins. Cell Death & Differentiation. 2008 Apr;15(4):635–41.

101. Wiesener MS, Jürgensen JS, Rosenberger C, Scholze CK, Hörstrup JH, Warnecke C, et al. Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs. FASEB J. 2003 Feb;17(2):271–3.

102. Semenza GL. Regulation of Mammalian O2 Homeostasis by Hypoxia-Inducible Factor 1. Annual Review of Cell and Developmental Biology. 1999;15(1):551–78.

103. Schönenberger MJ, Kovacs WJ. Hypoxia signaling pathways: modulators of oxygen-related organelles. Front Cell Dev Biol . 2015 Jul 21;3.

104. MacDonald BT, Tamai K, He X. Wnt/β-catenin signaling: components, mechanisms, and diseases. Developmental cell. 2009 Jul;17(1):9–26.

105. Gómez-Orte E, Sáenz-Narciso B, Moreno S, Cabello J. Multiple functions of the noncanonical Wnt pathway. Trends Genet. 2013 Sep;29(9):545–53.

106. Vladar EK, Antic D, Axelrod JD. Planar Cell Polarity Signaling: The Developing Cell's Compass. Cold Spring Harb Perspect Biol . 2009 Sep;1(3).

107. De A. Wnt/Ca2+ signaling pathway: a brief overview. Acta Biochim Biophys Sin (Shanghai). 2011 Oct 1;43(10):745–56.

108. Whyte JL, Smith AA, Helms JA. Wnt Signaling and Injury Repair. Cold Spring Harb Perspect Biol . 2012 Aug;4(8).

109. Mazumdar J, O'Brien WT, Johnson RS, LaManna JC, Chavez JC, Klein PS, et al. O2 regulates stem cells through Wnt/β-catenin signalling. Nat Cell Biol. 2010 Oct;12(10):1007–13.

110. Higgins DF, Kimura K, Bernhardt WM, Shrimanker N, Akai Y, Hohenstein B, et al. Hypoxia promotes fibrogenesis in vivo via HIF-1 stimulation of epithelial-to-mesenchymal transition. J Clin Invest. 2007 Dec 3;117(12):3810–20.

111. Chilosi M, Poletti V, Zamò A, Lestani M, Montagna L, Piccoli P, et al. Aberrant Wnt/β-Catenin Pathway Activation in Idiopathic Pulmonary Fibrosis. The American Journal of Pathology. 2003 May;162(5):1495–502.

112. Yang IV, Burch LH, Steele MP, Savov JD, Hollingsworth JW, McElvania-Tekippe E, et al. Gene Expression Profiling of Familial and Sporadic Interstitial Pneumonia. American Journal of Respiratory and Critical Care Medicine. 2007 Jan;175(1):45–54.

113. Akhmetshina A, Palumbo K, Dees C, Bergmann C, Venalis P, Zerr P, et al. Activation of canonical Wnt signalling is required for TGF-β-mediated fibrosis. Nature Communications. 2012 Mar;3:735.

114. Lam AP, Herazo-Maya JD, Sennello JA, Flozak AS, Russell S, Mutlu GM, et al. Wnt Coreceptor Lrp5 Is a Driver of Idiopathic Pulmonary Fibrosis. American Journal of Respiratory and Critical Care Medicine. 2014 Jun;190(2):185–95.

115. Flozak AS, Lam AP, Russell S, Jain M, Peled ON, Sheppard KA, et al. β-Catenin/T-cell Factor Signaling Is Activated during Lung Injury and Promotes the Survival and Migration of Alveolar Epithelial Cells. The Journal of Biological Chemistry. 2010 Jan;285(5):3157–67.

116. Aumiller V, Balsara N, Wilhelm J, Günther A, Königshoff M. WNT/β-Catenin Signaling Induces IL-1β Expression by Alveolar Epithelial Cells in Pulmonary Fibrosis. American Journal of Respiratory Cell and Molecular Biology. 2013 Mar;49(1):96–104.

117. Klee S, Lehmann M, Wagner DE, Baarsma HA, Königshoff M. WISP1 mediates IL-6-dependent proliferation in primary human lung fibroblasts. Scientific Reports. 2016 Feb;6:20547.

118. Nakamura T, Bornstein P. Matricellular Proteins☆. In: Reference Module in Biomedical Sciences. Elsevier; 2020.

119. Holbourn KP, Acharya KR, Perbal B. The CCN family of proteins: structure–function relationships. Trends in Biochemical Sciences. 2008 Oct;33(10):461–73.

120. Chen P-C, Cheng H-C, Yang S-F, Lin C-W, Tang C-H. The CCN Family Proteins: Modulators of Bone Development and Novel Targets in Bone-Associated Tumors . Vol. 2014, BioMed Research International. Hindawi; 2014. p. e437096.

121. Holbourn KP, Acharya KR, Perbal B. The CCN family of proteins: structure–function relationships. Trends in Biochemical Sciences. 2008 Oct;33(10):461–73.

122. Jun J-I, Lau LF. Taking Aim at the Extracellular Matrix: CCN Proteins as Emerging Therapeutic Targets. Nature reviews Drug discovery. 2011 Dec;10(12):945–63.

123. Jun J-I, Lau LF. The Matricellular Protein CCN1/CYR61 Induces Fibroblast Senescence and Restricts Fibrosis in Cutaneous Wound Healing. Nat Cell Biol. 2010 Jul;12(7):676–85.

124. Chen C-C, Young JL, Monzon RI, Chen N, Todorović V, Lau LF. Cytotoxicity of TNFα is regulated by integrin-mediated matrix signaling. EMBO J. 2007 Mar 7;26(5):1257–67.

125. Pi L, Robinson PM, Jorgensen M, Oh S-H, Brown AR, Weinreb PH, et al. Connective Tissue Growth Factor and Integrin αvβ6: a New Pair of Regulators Critical for Ductular Reaction and Biliary Fibrosis. Hepatology. 2015 Feb;61(2):678–91.

126. Lipson KE, Wong C, Teng Y, Spong S. CTGF is a central mediator of tissue remodeling and fibrosis and its inhibition can reverse the process of fibrosis. Fibrogenesis & Tissue Repair. 2012 Jun 6;5(1):S24.

127. Tsai C-C, Wu S-B, Kau H-C, Wei Y-H. Essential role of connective tissue growth factor (CTGF) in transforming growth factor-β1 (TGF-β1)-induced myofibroblast transdifferentiation from Graves' orbital fibroblasts. Scientific Reports. 2018 May 8;8(1):7276.

128. Wahab NA, Brinkman H, Mason RM. Uptake and intracellular transport of the connective tissue growth factor: a potential mode of action. Biochem J. 2001 Oct 1;359(Pt 1):89–97.

129. WISP1 protein expression summary - The Human Protein Atlas . [cited 2020 Dec 20]. Available from: https://www.proteinatlas.org/ENSG00000104415-WISP1

130. Malik AR, Liszewska E, Jaworski J. Matricellular proteins of the Cyr61/CTGF/NOV (CCN) family and the nervous system. Front Cell Neurosci . 2015;9.

131. Xu L, Corcoran RB, Welsh JW, Pennica D, Levine AJ. WISP-1 is a Wnt-1- and β-catenin-responsive oncogene. Genes & Development. 2000 Mar;14(5):585–95.

132. Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, et al. WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors. Proceedings of the National Academy of Sciences. 1998 Dec;95(25):14717–22.

133. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, et al. Large-scale mapping of human protein–protein interactions by mass spectrometry. Molecular Systems Biology. 2007 Jan;3(1):89-n/a.

134. Heise RL, Stober V, Cheluvaraju C, Hollingsworth JW, Garantziotis S. Mechanical Stretch Induces Epithelial-Mesenchymal Transition in Alveolar Epithelia via Hyaluronan Activation of Innate Immunity. J Biol Chem. 2011 May 20;286(20):17435–44.

135. Jiang D, Liang J, Fan J, Yu S, Chen S, Luo Y, et al. Regulation of lung injury and repair by Toll-like receptors and hyaluronan. Nature Medicine. 2005 Nov;11(11):1173–9.

136. Li H-H, Li Q, Liu P, Liu Y, Li J, Wasserloos K, et al. WNT1-Inducible Signaling Pathway Protein 1 Contributes to Ventilator-Induced Lung Injury. Am J Respir Cell Mol Biol. 2012 Oct;47(4):528–35.

137. Ding X, Wang X, Zhao X, Jin S, Tong Y, Ren H, et al. Rgd Peptides Protects Against Acute Lung Injury in Septic Mice Through Wisp1-integrin β6 Pathway Inhibition. Shock. 2015 Apr 1;43(4):352–60.

138. Chen Z, Ding X, Jin S, Pitt B, Zhang L, Billiar T, et al. WISP1-αvβ3 integrin signaling positively regulates TLR-triggered inflammation response in sepsis induced lung injury. Sci Rep. 2016 28;6:28841.

139. Yang M, Zhao X, Liu Y, Tian Y, Ran X, Jiang Y. A role for WNT1-inducible signaling protein-1 in airway remodeling in a rat asthma model. International Immunopharmacology. 2013;17(2):350–7.

140. Kaliyamurthi V, Venkatesan B, Valente A, Melby P, Nandish S, Reusch J, et al. WISP1, a Pro-mitogenic, Pro-survival Factor, Mediates Tumor Necrosis Factor- (TNF-)-stimulated Cardiac Fibroblast Proliferation but Inhibits TNF- -induced Cardiomyocyte Death. Vol. 284. 2009. 14414–14427 p.

141. Königshoff M, Kramer M, Balsara N, Wilhelm J, Amarie OV, Jahn A, et al. WNT1-inducible signaling protein–1 mediates pulmonary fibrosis in mice and is upregulated in humans with idiopathic pulmonary fibrosis. The Journal of Clinical Investigation. 2009 Apr;119(4):772–87.

142. Berschneider B, Ellwanger DC, Baarsma HA, Thiel C, Shimbori C, White ES, et al. miR-92a regulates TGF-β1-induced WISP1 expression in pulmonary fibrosis. The International Journal of Biochemistry & Cell Biology. 2014;53:432–41.

143. Ambler L. Expression and activity of wnt1-inducible signalling protein-1 in idiopathic pulmonary fibrosis. 2016. PhD Thesis.

144. Jian Y-C, Wang J-J, Dong S, Hu J-W, Hu L-J, Yang G-M, et al. Wnt-induced secreted protein 1/CCN4 in liver fibrosis both in vitro and in vivo. Clin Lab. 2014;60(1):29–35.

145. Yang X, Wang H, Tu Y, Li Y, Zou Y, Li G, et al. WNT1-inducible signaling protein-1 mediates TGF-β1-induced renal fibrosis in tubular epithelial cells and unilateral ureteral obstruction mouse models via autophagy. J Cell Physiol. 2020 Mar;235(3):2009–22.

146. Gurbuz I, Chiquet-Ehrismann R. CCN4/WISP1 (WNT1 inducible signaling pathway protein 1): A focus on its role in cancer. The International Journal of Biochemistry & Cell Biology. 2015 May 1;62:142–6.

147. Chiang K-C, Yeh C-N, Chung L-C, Feng T-H, Sun C-C, Chen M-F, et al. WNT-1 inducible signaling pathway protein-1 enhances growth and tumorigenesis in human breast cancer. Scientific Reports. 2015 Mar 3;5:8686.

148. Wu W, Liu X, Wei L, Li T, Zang Y, Qian Y, et al. Tp53 Mutation Inhibits Ubiquitination and Degradation of WISP1 via Down-Regulation of Siah1 in Pancreatic Carcinogenesis. Front

Pharmacol . 2018 Aug 3 [cited 2018 Nov 20];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6085464/

149. French DM, Kaul RJ, D'Souza AL, Crowley CW, Bao M, Frantz GD, et al. WISP-1 Is an Osteoblastic Regulator Expressed During Skeletal Development and Fracture Repair. Am J Pathol. 2004 Sep;165(3):855–67.

150. Maeda A, Ono M, Holmbeck K, Li L, Kilts TM, Kram V, et al. WNT1-induced Secreted Protein-1 (WISP1), a Novel Regulator of Bone Turnover and Wnt Signaling. J Biol Chem. 2015 May 29;290(22):14004–18.

151. Blom AB, Brockbank SM, van Lent PL, van Beuningen HM, Geurts J, Takahashi N, et al. Involvement of the Wnt signaling pathway in experimental and human osteoarthritis: prominent role of Wnt-induced signaling protein 1. Arthritis Rheum. 2009 Feb;60(2):501–12.

152. Ferrand N, Béreziat V, Moldes M, Zaoui M, Larsen AK, Sabbah M. WISP1/CCN4 inhibits adipocyte differentiation through repression of PPARγ activity. Scientific Reports. 2017 May 11;7(1):1749.

153. McGowan SE. The lipofibroblast: more than a lipid-storage depot. American Journal of Physiology-Lung Cellular and Molecular Physiology. 2019 May 1;316(5):L869–71.

154. Komori T. Signaling networks in RUNX2-dependent bone development. J Cell Biochem. 2011 Mar;112(3):750–5.

155. Meyers CA, Xu J, Asatrian G, Ding C, Shen J, Broderick K, et al. WISP-1 drives bone formation at the expense of fat formation in human perivascular stem cells. Scientific Reports. 2018 Oct 23;8(1):15618.

156. NCBI. WISP1 WNT1 inducible signaling pathway protein 1 [Homo sapiens (human)] . 2017. Available from: https://www.ncbi.nlm.nih.gov/gene/8840

157. Tanaka S, Sugimachi K, Saeki H, Kinoshita J, Ohga T, Shimada M, et al. A novel variant of WISP1 lacking a Von Willebrand type C module overexpressed in scirrhous gastric carcinoma. Oncogene. 2001;20(39):5525–32.

158. Tanaka S, Sugimachi K, Kameyama T, Maehara S, Shirabe K, Shimada M, et al. Human WISP1v, a member of the CCN family, is associated with invasive cholangiocarcinoma. Hepatology. 2003 May;37(5):1122–9.

159. Cervello M, Giannitrapani L, Labbozzetta M, Notarbartolo M, D'alessandro N, Lampiasi N, et al. Expression of WISPs and of Their Novel Alternative Variants in Human Hepatocellular Carcinoma Cells. Annals of the New York Academy of Sciences. 2004 Dec;1028(1):432–9.

160. Li Z, Chiao PJ, Evans DB. Homo sapiens WISP1c (WISP1) mRNA, complete cds; alternatively spliced . 2005. Available from: https://www.ncbi.nlm.nih.gov/nuccore/AY196486.1

161. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan;26(1):139–40.

162. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012 May;40(10):4288–97.

163. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015 Apr;43(7):e47–e47.

164. R Core Team. R: A Language and Environment for Statistical Computing . Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

165. Wickham H. ggplot2: Elegant Graphics for Data Analysis . Springer-Verlag New York; 2016. Available from: https://ggplot2.tidyverse.org

166. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009 Feb 3;10(1):48.

167. Eden E, Lipson D, Yogev S, Yakhini Z. Discovering Motifs in Ranked Lists of DNA Sequences. PLOS Computational Biology. 2007 Mar 23;3(3):e39.

168. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS. 2005 Oct 25;102(43):15545–50.

169. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.

170. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature Genetics. 2000 May 1;25:25–9.

171. Gene Ontology Resource: 20 years and still GOing strong | Nucleic Acids Research | Oxford Academic. 2019 Jan 8. 47(D1):D330-D338.

172. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 13;177(7):1888-1902.e21.

173. Bates D, Maechler M. Matrix: Sparse and Dense Matrix Classes and Methods . 2019. Available from: https://CRAN.R-project.org/package=Matrix

174. Maaten L van der, Hinton G. Visualizing Data using t-SNE. Journal of Machine Learning Research. 2008;9(86):2579–605.

175. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software. 2018 Sep 2;3(29):861.

176. Rusk N. Expanded CIBERSORTx. Nature Methods. 2019 Jul;16(7):577–577.

177. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015 May;12(5):453–7.

178. Buffa FM, Harris AL, West CM, Miller CJ. Large meta-analysis of multiple cancers reveals a common, compact and highly prognostic hypoxia metagene. Br J Cancer. 2010 Jan 19;102(2):428–35.

179. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 2013 Jan 16;14(1):7.

180. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature Biotechnology. 2014 Apr;32(4):381–6.

181. Molina-Arcas M, Hancock DC, Sheridan C, Kumar MS, Downward J. Coordinate direct input of both KRAS and IGF1 receptor to activation of PI 3-kinase in KRAS mutant lung cancer. Cancer Discov. 2013 May;3(5):548–63.

182. Coelho MA, de Carné Trécesson S, Rana S, Zecchin D, Moore C, Molina-Arcas M, et al. Oncogenic RAS Signaling Promotes Tumor Immunoresistance by Stabilizing PD-L1 mRNA. Immunity. 2017 Dec 19;47(6):1083-1099.e6.

183. Nishida T, Kubota S, Aoyama E, Janune D, Maeda A, Takigawa M. Effect of CCN2 on FGF2-Induced Proliferation and MMP9 and MMP13 Productions by Chondrocytes. Endocrinology. 2011 Nov 1;152(11):4232–41.

184. Meltzer EB, Barry WT, D'Amico TA, Davis RD, Lin SS, Onaitis MW, et al. Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. BMC Medical Genomics. 2011 Oct;4:70.

185. Nance T, Smith KS, Anaya V, Richardson R, Ho L, Pala M, et al. Transcriptome Analysis Reveals Differential Splicing Events in IPF Lung Tissue. Buratti E, editor. PLoS ONE. 2014 Mar;9(3):e92111–e92111.

186. Vukmirovic M, Kaminski N. Impact of Transcriptomics on Our Understanding of Pulmonary Fibrosis. Front Med (Lausanne). 2018 Apr 4;5.

187. Yang IV, Fingerlin TE, Evans CM, Schwarz MI, Schwartz DA. MUC5B and Idiopathic Pulmonary Fibrosis. Ann Am Thorac Soc. 2015 Nov;12(Suppl 2):S193–9.

188. Nakajima M, Kizawa H, Saitoh M, Kou I, Miyazono K, Ikegawa S. Mechanisms for Asporin Function and Regulation in Articular Cartilage. J Biol Chem. 2007 Feb 11;282(44):32185–92.

189. Niehrs C. Function and biological roles of the Dickkopf family of Wnt modulators. Oncogene. 2006 Dec 4;25(57):7469–81.

190. Gibb N, Lavery DL, Hoppler S. sfrp1 promotes cardiomyocyte differentiation in Xenopus via negative-feedback regulation of Wnt signalling. Development (Cambridge, England). 2013 Apr;140(7):1537–49.

191. Königshoff M, Kramer M, Balsara N, Wilhelm J, Amarie OV, Jahn A, et al. WNT1-inducible signaling protein–1 mediates pulmonary fibrosis in mice and is upregulated in humans with idiopathic pulmonary fibrosis. The Journal of Clinical Investigation. 2009 Apr;119(4):772–87.

192. Berschneider B, Königshoff M. WNT1 inducible signaling pathway protein 1 (WISP1): A novel mediator linking development and disease. The International Journal of Biochemistry & Cell Biology. 2011 Mar;43(3):306–9.

193. Meltzer EB, Noble PW. Idiopathic pulmonary fibrosis. Orphanet Journal of Rare Diseases. 2008 Mar 26;3(1):8.

194. Maiese K. Picking a bone with WISP1 (CCN4): new strategies against degenerative joint disease. Journal of translational science. 2016 Jan;1(3):83–5.

195. Chang W, Wei K, Jacobs SS, Upadhyay D, Weill D, Rosen GD. SPARC suppresses apoptosis of idiopathic pulmonary fibrosis fibroblasts through constitutive activation of beta-catenin. J Biol Chem. 2010 Mar 12;285(11):8196–206.

196. Liu Y, Bhat RA, Seestaller-Wehr LM, Fukayama S, Mangine A, Moran RA, et al. The Orphan Receptor Tyrosine Kinase Ror2 Promotes Osteoblast Differentiation and Enhances ex Vivo Bone Formation. Molecular Endocrinology. 2007 Feb 1;21(2):376–87.

197. Malinauskas T, Jones EY. Extracellular modulators of Wnt signalling. Curr Opin Struct Biol. 2014 Dec;29:77–84.

198. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Science Advances. 2020 Jul 1;6(28):eaba1983.

199. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, et al. Single Cell RNA-seq reveals ectopic and aberrant lung resident cell populations in Idiopathic Pulmonary Fibrosis. bioRxiv. 2019 Sep 6;759902.

200. Law RH, Zhang Q, McGowan S, Buckle AM, Silverman GA, Wong W, et al. An overview of the serpin superfamily. Genome Biol. 2006;7(5):216.

201. Hsieh H-H, Chen Y-C, Jhan J-R, Lin J-J. The serine protease inhibitor serpinB2 binds and stabilizes p21 in senescent cells. J Cell Sci. 2017 Oct 1;130(19):3272–81.

202. Okuda R, Aoshiba K, Matsushima H, Ogura T, Okudela K, Ohashi K. Cellular senescence and senescence-associated secretory phenotype: comparison of idiopathic pulmonary fibrosis, connective tissue disease-associated interstitial lung disease, and chronic obstructive pulmonary disease. J Thorac Dis. 2019 Mar;11(3):857–64.

203. Coppé J-P, Desprez P-Y, Krtolica A, Campisi J. The Senescence-Associated Secretory Phenotype: The Dark Side of Tumor Suppression. Annu Rev Pathol. 2010;5:99–118.

204. Chien Y, Scuoppo C, Wang X, Fang X, Balgley B, Bolden JE, et al. Control of the senescence-associated secretory phenotype by NF-κB promotes senescence and enhances chemosensitivity. Genes Dev. 2011 Oct 15;25(20):2125–36.

205. Litwiniuk M, Krejner A, Speyrer MS, Gauto AR, Grzela T. Hyaluronic Acid in Inflammation and Tissue Regeneration. Wounds. 2016 Mar;28(3):78–88.

206. Park J, Ivey MJ, Deana Y, Riggsbee KL, Sörensen E, Schwabl V, et al. The Tcf21 lineage constitutes the lung lipofibroblast population. Am J Physiol Lung Cell Mol Physiol. 2019 May 1;316(5):L872–85.

207. Vašák M. Advances in metallothionein structure and functions. Journal of Trace Elements in Medicine and Biology. 2005 Sep 19;19(1):13–7.

208. Ruttkay-Nedecky B, Nejdl L, Gumulec J, Zitka O, Masarik M, Eckschlager T, et al. The Role of Metallothionein in Oxidative Stress. Int J Mol Sci. 2013 Mar 15;14(3):6044–66.

209. Poss KD, Tonegawa S. Reduced stress defense in heme oxygenase 1-deficient cells. Proc Natl Acad Sci U S A. 1997 Sep 30;94(20):10925–30.

210. Liu X, Rowan SC, Liang J, Yao C, Huang G, Deng N, et al. Definition and Signatures of Lung Fibroblast Populations in Development and Fibrosis in Mice and Men. bioRxiv. 2020 Jul 15;2020.07.15.203141.

211. Aspal M, Zemans RL. Mechanisms of ATII-to-ATI Cell Differentiation during Lung Regeneration. Int J Mol Sci . 2020 Apr 30;21(9).

212. O'Dwyer DN, Moore BB. The Role of Periostin in Lung Fibrosis and Airway Remodeling. Cell Mol Life Sci. 2017 Dec;74(23):4305–14.

213. Okamoto M, Hoshino T, Kitasato Y, Sakazaki Y, Kawayama T, Fujimoto K, et al. Periostin, a matrix protein, is a novel biomarker for idiopathic interstitial pneumonias. Eur Respir J. 2011 May;37(5):1119–27.

214. Zhang X, Chen X, Liu J, Dong X, Jin Y, Tian Y, et al. Knockdown of WISP1 inhibit proliferation and induce apoptosis in ALL Jurkat cells. Int J Clin Exp Pathol. 2015 Nov 1;8(11):15489–96.

215. Brereton C, Jones MG, Zhou Y, Conforti F, Bell J, Davies DE. Non-canonical HIF pathway activation dysregulates collagen structure-function in human lung fibrosis. Unpublished Manuscript.

216. Kawano Y, Kypta R. Secreted antagonists of the Wnt signalling pathway. Journal of Cell Science. 2003 Jul 1;116(13):2627–34.

217. Upagupta C, Shimbori C, Alsilmi R, Kolb M. Matrix abnormalities in pulmonary fibrosis. European Respiratory Review . 2018 Jun 30;27(148).

218. Fernandez IE, Eickelberg O. The Impact of TGF-β on Lung Fibrosis. Proceedings of the American Thoracic Society. 2012 Jul;9(3):111–6.

219. Biernacka A, Dobaczewski M, Frangogiannis NG. TGF-β signaling in fibrosis. Growth Factors. 2011 Oct;29(5):196–202.

220. Komiya Y, Habas R. Wnt signal transduction pathways. Organogenesis. 2008;4(2):68–75.

221. Abedini A, Sayed C, Carter LE, Boerboom D, Vanderhyden BC. Non-canonical WNT5a regulates Epithelial-to-Mesenchymal Transition in the mouse ovarian surface epithelium. Scientific Reports. 2020 Jun 16;10(1):9695.

222. Shimizu H, Julius MA, Giarré M, Zheng Z, Brown AM, Kitajewski J. Transformation by Wnt family proteins correlates with regulation of beta-catenin. Cell Growth Differ. 1997 Dec;8(12):1349–58.

223. Vuga LJ, Ben-Yehudah A, Kovkarova-Naumovski E, Oriss T, Gibson KF, Feghali-Bostwick C, et al. WNT5A Is a Regulator of Fibroblast Proliferation and Resistance to Apoptosis. Am J Respir Cell Mol Biol. 2009 Nov;41(5):583–9.

224. Newman DR, Sills WS, Hanrahan K, Ziegler A, Tidd KM, Cook E, et al. Expression of WNT5A in Idiopathic Pulmonary Fibrosis and Its Control by TGF-β and WNT7B in Human Lung Fibroblasts. J Histochem Cytochem. 2016 Feb;64(2):99–111.

225. Sánchez-Elsner T, Ramírez JR, Rodriguez-Sanz F, Varela E, Bernabéu C, Botella LM. A Cross-Talk Between Hypoxia and TGF-β Orchestrates Erythropoietin Gene Regulation Through SP1 and Smads. Journal of Molecular Biology. 2004 Feb 6;336(1):9–24.

226. Semenza GL. Oxygen sensing, hypoxia-inducible factors, and disease pathophysiology. Annu Rev Pathol. 2014;9:47–71.

227. Gorres KL, Raines RT. Prolyl 4-hydroxylase. Crit Rev Biochem Mol Biol. 2010 Apr;45(2):106–24.

228. Farmer SR. Transcriptional control of adipocyte formation. Cell Metab. 2006 Oct;4(4):263–73.

229. Komori T. Signaling networks in RUNX2-dependent bone development. Journal of Cellular Biochemistry. 2011;112(3):750–5.

230. Yanagita T, Kubota S, Kawaki H, Kawata K, Kondo S, Takano-Yamamoto T, et al. Expression and physiological role of CCN4/Wnt-induced secreted protein 1 mRNA splicing variants in chondrocytes. FEBS Journal. 2007 Apr;274(7):1655–65.

231. Moon J-O, Welch TP, Gonzalez FJ, Copple BL. Reduced liver fibrosis in hypoxia-inducible factor-1α-deficient mice. Am J Physiol Gastrointest Liver Physiol. 2009 Mar;296(3):G582–92.

232. Masson N, Singleton RS, Sekirnik R, Trudgian DC, Ambrose LJ, Miranda MX, et al. The FIH hydroxylase is a cellular peroxide sensor that modulates HIF transcriptional activity. EMBO Rep. 2012 Mar;13(3):251–7.

233. Cheresh P, Kim S-J, Tulasiram S, Kamp DW. Oxidative Stress and Pulmonary Fibrosis. Biochim Biophys Acta. 2013 Jul;1832(7):1028–40.

234. Murahovschi V, Pivovarova O, Ilkavets I, Dmitrieva RM, Döcke S, Keyhani-Nejad F, et al. WISP1 Is a Novel Adipokine Linked to Inflammation in Obesity. Diabetes. 2015 Mar 1;64(3):856–66.

235. Cernea M, Tang W, Guan H, Yang K. Wisp1 mediates Bmp3-stimulated mesenchymal stem cell proliferation. Journal of Molecular Endocrinology. 2016 Jan 1;56(1):39–46.

236. Showalter AE, Martini AC, Nierenberg D, Hosang K, Fahmi NA, Gopalan P, et al. Investigating Chaperonin-Containing TCP-1 subunit 2 as an essential component of the chaperonin complex for tumorigenesis. Scientific Reports. 2020 Jan 21;10(1):798.

237. Xiao L, Du Y, Shen Y, He Y, Zhao H, Li Z. TGF-beta 1 induced fibroblast proliferation is mediated by the FGF-2/ERK pathway. Front Biosci (Landmark Ed). 2012 Jun 1;17:2667–74.

238. Raballo R, Rhee J, Lyn-Cook R, Leckman JF, Schwartz ML, Vaccarino FM. Basic Fibroblast Growth Factor (Fgf2) Is Necessary for Cell Proliferation and Neurogenesis in the Developing Cerebral Cortex. J Neurosci. 2000 Jul 1;20(13):5012–23.

239. Merkwirth C, Langer T. Prohibitin function within mitochondria: Essential roles for cell proliferation and cristae morphogenesis. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2009 Jan 1;1793(1):27–32.

240. Abu-Hamad S, Arbel N, Calo D, Arzoine L, Israelson A, Keinan N, et al. The VDAC1 N-terminus is essential both for apoptosis and the protective effect of anti-apoptotic proteins. J Cell Sci. 2009 Jun 1;122(Pt 11):1906–16.

241. Shoshan-Barmatz V, Maldonado EN, Krelin Y. VDAC1 at the crossroads of cell metabolism, apoptosis and cell stress. Cell Stress. 1(1):11–36.

242. Shao S, Fang H, Duan L, Ye X, Rao S, Han J, et al. Lysyl hydroxylase 3 increases collagen deposition and promotes pulmonary fibrosis by activating TGFβ1/Smad3 and Wnt/β-catenin pathways. Arch Med Sci. 2020;16(2):436–45.

243. Benanti JA, Williams DK, Robinson KL, Ozer HL, Galloway DA. Induction of Extracellular Matrix-Remodeling Genes by the Senescence-Associated Protein APA-1. Mol Cell Biol. 2002 Nov;22(21):7385–97.

244. Dupree MA, Pollack SR, Levine EM, Laurencin CT. Fibroblast Growth Factor 2 Induced Proliferation in Osteoblasts and Bone Marrow Stromal Cells: A Whole Cell Model. Biophys J. 2006 Oct 15;91(8):3097–112.

245. Bugler B, Amalric F, Prats H. Alternative initiation of translation determines cytoplasmic or nuclear localization of basic fibroblast growth factor. Mol Cell Biol. 1991 Jan;11(1):573–7.

246. Chen C-C, Lau LF. Functions and mechanisms of action of CCN matricellular proteins. The International Journal of Biochemistry & Cell Biology. 2009 Apr 1;41(4):771–83.

247. Hopp TP, Prickett KS, Price VL, Libby RT, March CJ, Pat Cerretti D, et al. A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. Bio/Technology. 1988 Oct;6(10):1204–10.

248. Beilina A, Bonet-Ponce L, Kumaran R, Kordich JJ, Ishida M, Mamais A, et al. The Parkinson's Disease Protein LRRK2 Interacts with the GARP Complex to Promote Retrograde Transport to the trans-Golgi Network. Cell Rep. 2020 May 5;31(5):107614.

249. Vijen S, Hawes C, Runions J, Russell RGG, Wordsworth BP, Carr AJ, et al. Differences in intracellular localisation of ANKH mutants that relate to mechanisms of calcium pyrophosphate deposition disease and craniometaphyseal dysplasia. Sci Rep . 2020 May 4;10.

250. Miura K. An Overview of Current Methods to Confirm Protein-Protein Interactions. Protein Pept Lett. 2018 Aug;25(8):728–33.

251. Hecht ES, Scigelova M, Eliuk S, Makarov A. Fundamentals and Advances of Orbitrap Mass Spectrometry. In: Encyclopedia of Analytical Chemistry. American Cancer Society; 2019. p. 1–40.

252. Freund A, Zhong FL, Venteicher AS, Meng Z, Veenstra TD, Frydman J, et al. Proteostatic control of telomerase function through TRiC-mediated folding of TCAB1. Cell. 2014 Dec 4;159(6):1389–403.

253. Venteicher AS, Artandi SE. TCAB1: Driving telomerase to Cajal bodies. Cell Cycle. 2009 May 1;8(9):1329–31.

254. Pear WS, Nolan GP, Scott ML, Baltimore D. Production of high-titer helper-free retroviruses by transient transfection. Proc Natl Acad Sci U S A. 1993 Sep 15;90(18):8392–6.

255. Bilgili H, Białas AJ, Górski P, Piotrowski WJ. Telomere Abnormalities in the Pathobiology of Idiopathic Pulmonary Fibrosis. J Clin Med. 2019 Aug 16.

256. Schittek B. The Multiple Facets of Dermcidin in Cell Survival and Host Defense. JIN. 2012;4(4):349–60.

257. Porter D, Weremowicz S, Chin K, Seth P, Keshaviah A, Lahti-Domenici J, et al. A neural survival factor is a candidate oncogene in breast cancer. Proc Natl Acad Sci U S A. 2003 Sep 16;100(19):10931–6.

258. Cunningham TJ, Hodge L, Speicher D, Reim D, Tyler-Polsz C, Levitt P, et al. Identification of a Survival-Promoting Peptide in Medium Conditioned by Oxidatively Stressed Cell Lines of Nervous System Origin. J Neurosci. 1998 Sep 15;18(18):7047–60.

259. Stewart GD, Lowrie AG, Riddick ACP, Fearon KCH, Habib FK, Ross JA. Dermcidin expression confers a survival advantage in prostate cancer cells subjected to oxidative stress or hypoxia. The Prostate. 2007;67(12):1308–17.

260. Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods. 2013 Aug;10(8):730–6.

261. Merkwirth C, Dargazanli S, Tatsuta T, Geimer S, Löwer B, Wunderlich FT, et al. Prohibitins control cell proliferation and apoptosis by regulating OPA1-dependent cristae morphogenesis in mitochondria. Genes Dev. 2008 Feb 15;22(4):476–88.

262. Nijtmans LGJ, Artal SM, Grivell LA, Coates PJ. The mitochondrial PHB complex: roles in mitochondrial respiratory complex assembly, ageing and degenerative disease. Cell Mol Life Sci. 2002 Jan;59(1):143–55.

263. VDAC1 voltage dependent anion channel 1 [Homo sapiens (human)] - Gene - NCBI . [cited 2020 Dec 12]. Available from: https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=7416

264. Camara AKS, Zhou Y, Wen P-C, Tajkhorshid E, Kwok W-M. Mitochondrial VDAC1: A Key Gatekeeper as Potential Therapeutic Target. Front Physiol. 2017;8. Available from:

265. Bernard K, Logsdon NJ, Ravi S, Xie N, Persons BP, Rangarajan S, et al. Metabolic Reprogramming Is Required for Myofibroblast Contractility and Differentiation. J Biol Chem. 2015 Oct 16;290(42):25427–38.

266. Brahimi-Horn MC, Giuliano S, Saland E, Lacas-Gervais S, Sheiko T, Pelletier J, et al. Knockout of Vdac1 activates hypoxia-inducible factor through reactive oxygen species generation and induces tumor growth by promoting metabolic reprogramming and inflammation. Cancer Metab. 2015 Aug 26;3.

267. Adamis AP. CHAPTER 4 - Ocular angiogenesis: vascular endothelial growth factor and other factors. In: Nguyen QD, Rodrigues EB, Farah ME, Mieler WF, editors. Retinal Pharmacotherapy. Edinburgh: W.B. Saunders; 2010. p. 23–36.

268. Arif T, Vasilkovsky L, Refaely Y, Konson A, Shoshan-Barmatz V. Silencing VDAC1 Expression by siRNA Inhibits Cancer Cell Proliferation and Tumor Growth In Vivo. Molecular Therapy - Nucleic Acids. 2014 Jan 1;3.

269. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The Technology and Biology of Single-Cell RNA Sequencing. Molecular Cell. 2015 May 21;58(4):610–20.

270. Dimova DK, Dyson NJ. The E2F transcriptional network: old acquaintances with new faces. Oncogene. 2005 Apr;24(17):2810–26.

271. Atkinson JJ, Senior RM. Matrix Metalloproteinase-9 in Lung Remodeling. Am J Respir Cell Mol Biol. 2003 Jan 1;28(1):12–24.

272. Mushiroda T, Wattanapokayakit S, Takahashi A, Nukiwa T, Kudoh S, Ogura T, et al. A genome-wide association study identifies an association of a common variant in TERT with susceptibility to idiopathic pulmonary fibrosis. J Med Genet. 2008 Oct;45(10):654–6.

273. Kobayashi T, Kim H, Liu X, Sugiura H, Kohyama T, Fang Q, et al. Matrix metalloproteinase-9 activates TGF-β and stimulates fibroblast contraction of collagen gels. Am J Physiol Lung Cell Mol Physiol. 2014 Jun 1;306(11):L1006–15.

274. Ono M, Inkson CA, Kilts TM, Young MF. WISP-1/CCN4 regulates osteogenesis by enhancing BMP-2 activity. J Bone Miner Res. 2011 Jan;26(1):193–208.

275. Yoshioka Y, Ono M, Maeda A, Kilts TM, Hara ES, Khattab H, et al. CCN4/WISP-1 positively regulates chondrogenesis by controlling TGF-β3 function. Bone. 2016 Feb;83:162–70.

276. Rosset EM, Bradshaw AD. SPARC/osteonectin in mineralized tissue. Matrix Biol. 2016 Jul;52–54:78–87.

277. Ornitz DM, Itoh N. The Fibroblast Growth Factor signaling pathway. Wiley Interdiscip Rev Dev Biol. 2015 May;4(3):215–66.

278. Shi H-X, Lin C, Lin B-B, Wang Z-G, Zhang H-Y, Wu F-Z, et al. The Anti-Scar Effects of Basic Fibroblast Growth Factor on the Wound Repair In Vitro and In Vivo. PLoS One . 2013 Apr 2;8(4).

279. Eto H, Suga H, Aoi N, Kato H, Doi K, Kuno S, et al. Therapeutic potential of fibroblast growth factor-2 for hypertrophic scars: upregulation of MMP-1 and HGF expression. Lab Invest. 2012 Feb;92(2):214–23.

280. Dolivo DM, Larson SA, Dominko T. FGF2-mediated attenuation of myofibroblast activation is modulated by distinct MAPK signaling pathways in human dermal fibroblasts. J Dermatol Sci. 2017 Dec;88(3):339–48.

281. Wang F, Yang L, Shi L, Li Q, Zhang G, Wu J, et al. Nuclear translocation of fibroblast growth factor-2 (FGF2) is regulated by Karyopherin-β2 and Ran GTPase in human glioblastoma cells. Oncotarget. 2015 May 27;6(25):21468–78.

282. Chlebova K, Bryja V, Dvorak P, Kozubik A, Wilcox WR, Krejci P. High molecular weight FGF2: the biology of a nuclear growth factor. Cell Mol Life Sci. 2009 Jan;66(2):225–35.

283. Dini G, Funghini S, Witort E, Magnelli L, Fanti E, Rifkin DB, et al. Overexpression of the 18 kDa and 22/24 kDa FGF-2 isoforms results in differential drug resistance and amplification potential. J Cell Physiol. 2002 Oct;193(1):64–72.

284. Yue X, Shan B, Lasky JA. TGF-β: Titan of Lung Fibrogenesis. Curr Enzym Inhib. 2010 Jul 1;6(2).

285. Mingyuan X, Qianqian P, Shengquan X, Chenyi Y, Rui L, Yichen S, et al. Hypoxia-inducible factor-1α activates transforming growth factor-β1/Smad signaling and increases collagen deposition in dermal fibroblasts. Oncotarget. 2017 Dec 14;9(3):3188–97.

286. Furuta C, Miyamoto T, Takagi T, Noguchi Y, Kaneko J, Itoh S, et al. Transforming growth factor-β signaling enhancement by long-term exposure to hypoxia in a tumor microenvironment composed of Lewis lung carcinoma cells. Cancer Sci. 2015 Nov;106(11):1524–33.

287. Fois AG, Paliogiannis P, Sotgia S, Mangoni AA, Zinellu E, Pirina P, et al. Evaluation of oxidative stress biomarkers in idiopathic pulmonary fibrosis and therapeutic applications: a systematic review. Respir Res . 2018;19.